HAL
open science

On the Existence of Monge Maps for the Gromov-Wasserstein Problem

Théo Dumont, Théo Lacombe, François-Xavier Vialard

To cite this version:

Théo Dumont, Théo Lacombe, François-Xavier Vialard. On the Existence of Monge Maps for the Gromov-Wasserstein Problem. Foundations of Computational Mathematics, In press, 10.1007/s10208-024-09643-0 . hal-03818500v2

HAL Id: hal-03818500

https://hal.science/hal-03818500v2

Submitted on 6 Dec 2022 (v2), last revised 22 May 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE EXISTENCE OF MONGE MAPS FOR THE GROMOV–WASSERSTEIN PROBLEM

Théo Dumont, Théo Lacombe and François-Xavier Vialard*

Laboratoire d'Informatique Gaspard Monge,
Univ. Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

Abstract

In this work, we study the structure of minimizers of the quadratic Gromov–Wasserstein (GW) problem on Euclidean spaces for two different costs. The first one is the scalar product for which we prove that it is always possible to find optimizers as Monge maps and we detail the structure of such optimal maps. The second cost is the squared Euclidean distance for which we show that the worst case scenario is the existence of a bi-map structure. Both results are direct and indirect consequences of an existence result of optimal maps in the standard optimal transportation problem for costs that are defined by submersions. In dimension one for the squared Euclidean distance, we show numerical evidence for a negative answer to the existence of a Monge map under the conditions of Brenier's theorem, suggesting that our result cannot be improved in general. In addition, we show that a monotone map is optimal in some non-symmetric situations, thereby giving insight on why such a map often appears to be optimal in numerical experiments.

Contents

1 Introduction 2
1.1 The Gromov–Wasserstein problem 3
1.1.1 Formulation 3
1.1.2 Relation with the optimal transportation problem: a tight bi-convex relaxation 4
1.2 Related work 5
1.2.1 Monge maps for the OT problem 5
1.2.2 Monge maps for the GW problem 7
1.3 Outline, contributions and perspectives 8
2 Existence of Monge maps for fiber-invariant costs 8
2.1 Statement of the results 8
2.2 Proof of Theorem 2 11
2.3 Proof of Theorem 3 13
3 Applications to the quadratic and inner-product GW problems 14
3.1 The inner-product cost 15
3.2 The quadratic cost 16
3.3 Complementary study of the quadratic cost in the one-dimensional case 18
3.3.1 Adversarial computation of non-monotone optimal correspondence plans 20
3.3.2 Empirical instability of the optimality of monotone rearrangements 22
3.3.3 A positive result for measures with two components 24

*Email addresses: dumont.theo@protonmail.com, theo.lacombe@univ-eiffel.fr, francois-xavier.vialard@u-pem.fr.

1 Introduction

Finding correspondences between objects that do not live in the same metric space is a problem of fundamental interest both in applications and in theory, in very different fields such as computer vision and shape analysis [Ume88, MS04, MS05, BBM05, Mém07, Mém11], mathematics [Stu12], biology [DSS+20] and machine learning [RVFC20, AMJ18]. Graph matching [ZDlT15] is a prominent example of such a problem. A usual situation is when the objects of interest are metric spaces themselves, whose comparison is of practical and theoretical importance [GKPS99]; in this context, the Gromov–Hausdorff distance has been used in different settings [Mém08, MSW21] and its relaxation to L^p spaces, the Gromov–Wasserstein distance, has been explored in [GKPS99, Chap. 3 1/2] as well as in [Stu06, Mém07, Mém08]. Following this line of research, comparing metric measure spaces (*i.e.* metric spaces endowed with probability measures) using Wasserstein-type distances has attracted a lot of interest [Mém11, Stu12]. Follow-up works over the past decade include for instance [SVP21, DPM22] which propose extensions to the case of metric spaces with positive measures (*i.e.* whose total volume is not normalized to 1). The Gromov–Wasserstein distance and its extensions are applied successfully in machine learning [GJB19, BAMKJ19, XLH+20] and biology [DSC+22].

Following [Mém07], the Gromov–Wasserstein approach consists in seeking for a map between the two objects of interest that is of low distortion. In the case of measures that live in a metric space, this distortion is measured in terms of distances. To make it well-posed and symmetric, the problem is relaxed to a superposition of deterministic maps, such as in optimal transport where they are called plans or couplings [San15]. However, in stark contrast to optimal transport which is a linear programming problem, the formulation of the problems mentioned above falls in the class of quadratic assignment problems [KB57], which are computationally harder. As a consequence, it is not surprising that fewer results are available in the literature, yet some particular instances can be computed in polynomial time [MSW21]. Our interest in this paper is the understanding of the structure of optimal plans for this problem, and in particular when their support is the graph of a map. This problem has been put forward by Sturm in [Stu12, Challenge 3.6] and by Mémoli and Needham in [MN22, Question 2.14].

In optimal transport, the fact that the optimization can under mild conditions be reduced to the space of maps has been developed a lot since Brenier's work [Bre87] and further generalized by McCann [McC01]. Brenier's result essentially states that for the quadratic cost in Euclidean spaces, the optimal coupling is a map which is given by the gradient of a convex function. Such results on the structure of optimal plans/maps are of great interest in order to reduce the optimization set [MTOL20].

In optimal transport, the existence of optimal maps heavily depends on the choice of the cost functions. This is also the case for the Gromov–Wasserstein problem. In this work, we address the question of optimal maps for the Gromov–Wasserstein problem for two particular costs in Euclidean spaces. The first one is when the distortion is measured in terms of the scalar product; we show the existence of optimal maps and we detail their structure, under conditions similar to Brenier's theorem. The second case is the quadratic squared distance for which the problem seems to have less structure; we show that the worst case (but probably generic) situation is that optimal plans can always be chosen to be supported on the union of two graphs. We also study the one-dimensional case, which has attracted recent attention [Vay20, BHS22]. Indeed, in the latter article, a counter-example is given to the fact that the monotone (non-decreasing or non-increasing) mapping is optimal in the discrete case. We improve on these results in two directions. First by showing that this property is true under some specific conditions on the measures and second by providing numerical evidence for a counter-example to the existence of optimal maps between an arbitrary measure and a measure which is absolutely continuous w.r.t. the Lebesgue measure. We refer the reader to Section 1.3 for a detailed account of our contributions, while the background and state-of-the-art are presented respectively in Sections 1.1 and 1.2.

1.1 The Gromov–Wasserstein problem

1.1.1 Formulation. The Gromov–Wasserstein (GW) problem, initially introduced in [Mém07], can be seen as an extension of the Gromov–Hausdorff distance [GKPS99], to the context of (probability) measure spaces $(\mathcal{X}, \mu)$ equipped with a cost function $c_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (typically, $c_{\mathcal{X}}$ can be a distance on $\mathcal{X}$). Given $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ equipped with costs $c_{\mathcal{X}}, c_{\mathcal{Y}}$ respectively, and random variables $X, X' \sim \mu$ and $Y, Y' \sim \nu$, the GW problem seeks a correspondence (*i.e.* a joint law) between μ and ν that would make the distribution $c_{\mathcal{X}}(X, X')$ as close as possible to $c_{\mathcal{Y}}(Y, Y')$, in a L^p sense. Formally, it reads

Definition 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and $p \geq 1$. Given two probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, two continuous functions $c_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $c_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, the p-*Gromov–Wasserstein problem *aims at finding*

$$\mathrm{GW}_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathcal{X}\times\mathcal{Y}} \int_{\mathcal{X}\times\mathcal{Y}} |c_{\mathcal{X}}(x, x') - c_{\mathcal{Y}}(y, y')|^p \, \mathrm{d}\pi(x, y) \, \mathrm{d}\pi(x', y') \right)^{1/p}, \tag{GW}$$

where $\Pi(\mu, \nu)$ denotes the subset of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ of probability measures that admit μ (resp. ν) as first (resp. second) marginal. Any $\pi^\star$ minimizing (GW) is said to be an optimal correspondence plan *between μ and ν. If there exists a measurable map $T : \mathcal{X} \to \mathcal{Y}$ satisfying $T_\# \mu(A) \triangleq \mu(T^{-1}(A)) = \nu(A)$ for all Borel set A and such that $\pi^\star$ can be written as $\pi^\star = (\mathrm{id}, T)_\# \mu$, then T is said to be an* optimal correspondence map*, or a* Monge map *between μ and ν.*

While the existence of optimal correspondence plans holds under mild assumptions by compactness arguments as long as the above minimum is not $+\infty$, much less is known about the existence of optimal correspondence maps, even in simple cases.

In this work, we will consider two specific instances of this problem, both assuming that $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^d$ for two integers $n \geq d$ and using $p = 2$:

(i) the *inner product case* (considered in [AMJ18] and [Vay20]), where $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ are the inner products on $\mathbb{R}^n$ and $\mathbb{R}^d$ respectively (both denoted by $\langle \cdot, \cdot \rangle$):

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \int_{\mathcal{X}\times\mathcal{Y}} |\langle x, x' \rangle - \langle y, y' \rangle|^2 \, \mathrm{d}\pi(x, y) \, \mathrm{d}\pi(x', y'), \tag{GW-IP}$$

which essentially compares distribution of angles in $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$;

(ii) the *quadratic case* (considered in [Stu12] and [Vay20]), where $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ are the squared Euclidean distance on $\mathbb{R}^n$ and $\mathbb{R}^d$ respectively:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \int_{\mathcal{X}\times\mathcal{Y}} \left| |x - x'|^2 - |y - y'|^2 \right|^2 \, \mathrm{d}\pi(x, y) \, \mathrm{d}\pi(x', y'), \tag{GW-Q}$$

where by $|\cdot|$ we mean $\|\cdot\|_2$ to alleviate notations in the paper.

This second choice and more generally $c_{\mathcal{X}} = d_{\mathcal{X}}^q$ and $c_{\mathcal{Y}} = d_{\mathcal{Y}}^q$ with $q \geq 1$ is standard as we have the following property: if $\mathrm{GW}_p(\mu, \nu) = 0$, the metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu)$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$ are *strongly isomorphic*, that is there exists an isometry $\varphi : (\mathcal{X}, d_{\mathcal{X}}) \to (\mathcal{Y}, d_{\mathcal{Y}})$ such that $\varphi_\# \mu = \nu$, see [Stu12]. A subcase of this problem is given when $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ and $\nu = \frac{1}{N} \sum_{j=1}^N \delta_{y_j}$ are uniform probability distributions supported on N points each. In this scenario, optimal correspondence plans π can be chosen as permutations σ of $\{1, \ldots, N\}$ (see [ML18, Thm. 1] and [Vay20, Thm. 4.1.2]), and the problem optimizes over the set of such permutations $\mathfrak{S}_N$,

$$\min_{\sigma \in \mathfrak{S}_N} \sum_{i,j} \left| |x_i - x_j|^2 - |y_{\sigma(i)} - y_{\sigma(j)}|^2 \right|^2, \tag{QAP}$$

which is a particular case of the *Quadratic Assignment Problem* first introduced in [KB57].

1.1.2 Relation with the optimal transportation problem: a tight bi-convex relaxation. Let us first recall the formulation of the Optimal Transportation (OT) problem, also known as the Kantorovich problem, that will play an extensive role in this work.

Definition 2 (Kantorovich problem)**.** *Given two probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ and a cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$, we consider the problem*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\pi(x,y)\,. \tag{OT}$$

A transport plan $\pi \in \Pi(\mu,\nu)$ *realizing* (OT) *is called an* optimal transport plan, *or* optimal coupling. *Whenever it can be written as* $(\mathrm{id}, T)_\#\mu$ *for some map* $T : \mathcal{X} \to \mathcal{Y}$, *$T$ is said to be an* optimal transport map, *or a* Monge map *between μ and ν for the cost c.*

The minimization problem in (GW) can be interpreted as the minimization of the map $\pi \mapsto F(\pi,\pi) \triangleq \iint k\,\mathrm{d}\pi\otimes\pi$ where $k((x,y),(x',y')) = |c_{\mathcal{X}}(x,x') - c_{\mathcal{Y}}(y,y')|^2$, and F is thus a symmetric bilinear map. By first order condition, if $\pi^\star$ minimizes (GW), then it also minimizes $\pi \mapsto 2F(\pi,\pi^\star)$. If we let $C_{\pi^\star}(x,y) = \int_{\mathcal{X}\times\mathcal{Y}} k((x,y),(x',y'))\,\mathrm{d}\pi^\star(x',y')$, we obtain the linear problem

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} C_{\pi^\star}(x,y)\,\mathrm{d}\pi(x,y)\,, \tag{1}$$

which is nothing but the (OT) problem induced by the cost $C_{\pi^\star}$ on $\mathcal{X}\times\mathcal{Y}$. Therefore, we obtain that any optimal *correspondence* plan for (GW) with costs $c_{\mathcal{X}}, c_{\mathcal{Y}}$ must be an optimal *transportation* plan for (OT) with cost $C_{\pi^\star}$. A crucial point, proved in [SVP21, Thm. 3] as a generalization of [Kon76], is that if k is symmetric negative on the set of (signed) measures on $\mathcal{X}\times\mathcal{Y}$ with null marginals, that is $\int k\,\mathrm{d}\alpha\otimes\alpha \le 0$ for all such α, then the converse implication holds: any solution $\gamma^\star \in \Pi(\mu,\nu)$ of the OT problem with cost $C_{\pi^\star}$ is also a solution of the GW problem, that is

$$F(\pi^\star,\pi^\star) = F(\gamma^\star,\gamma^\star) = F(\pi^\star,\gamma^\star)\,. \tag{2}$$

Such a result is useful in this article to derive theoretical properties on the minimizers but also for the alternate minimization algorithms that were first proposed in [Mém07]. The question of the existence and structure of optimal maps has been extensively studied in optimal transportation, see Section 1.2. Since in this case the solutions of (GW) are in correspondence with the solutions of an OT problem, the tools and knowledge from optimal transportation can be used to derive existence and structure of optimal maps.

In particular, this holds for our two problems of interest (GW-Q) and (GW-IP): if α denotes a finite signed measure on $\mathcal{X}\times\mathcal{Y} \subset \mathbb{R}^n \times \mathbb{R}^d$ with null marginals, observe that

$$\begin{aligned}
&\iint \left||x-x'|^2 - |y-y'|^2\right|^2 \mathrm{d}\alpha(x,y)\,\mathrm{d}\alpha(x',y') \\
&\quad= \underbrace{\int |x-x'|^2\,\mathrm{d}\alpha\otimes\alpha}_{=0} + \underbrace{\int |y-y'|^2\,\mathrm{d}\alpha\otimes\alpha}_{=0} - 2\int |x-x'|^2|y-y'|^2\,\mathrm{d}\alpha\otimes\alpha \\
&\quad= -2\int (|x|^2 - 2\langle x,x'\rangle + |x|^2)(|y|^2 - 2\langle y,y'\rangle + |y'|^2)\,\mathrm{d}\alpha\otimes\alpha\,.
\end{aligned}$$

Developing the remaining factor involves nine terms, but given that α has null marginals (in particular, null mass), we obtain that $\int |x|^2|y|^2\,\mathrm{d}\alpha\otimes\alpha = 0$ (and similarly for the terms involving $|x'|^2|y'|^2$, $|x|^2|y'|^2$ and $|x'|^2|y|^2$), and also that $\int |x|^2\langle y,y'\rangle\,\mathrm{d}\alpha\otimes\alpha = 0$ (and similarly for the other terms). Eventually, the only remaining term is

$$-8\int \langle x,x'\rangle\langle y,y'\rangle\,\mathrm{d}\alpha\otimes\alpha = -8\left\|\int x\otimes y\,\mathrm{d}\alpha(x,y)\right\|_F^2 \le 0\,,$$

where $x \otimes y \in \mathbb{R}^{n\times d}$ is the matrix $(x_i y_j)_{i,j}$, where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_d)$, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix; the same holds for non-finite α by density. The negativity of this term ensures that solutions of (GW-Q) are exactly the solutions of an OT problem. Computations for (GW-IP) are similar—actually, they immediately boil down to the same last two equations. More generally, when one considers a cost such as $(d_{\mathcal{X}}(x,x') - d_{\mathcal{Y}}(y,y'))^2$, by expanding the square, the only term that matters in the optimization is $-2d_{\mathcal{X}}(x,x')d_{\mathcal{Y}}(y,y')$. Whenever it is possible to write both distances $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ as squared distances in Hilbert spaces, namely $d_{\mathcal{X}}(x,x') = \|\varphi(x) - \varphi(x')\|^2_{H_{\mathcal{X}}}$ and $d_{\mathcal{Y}}(y,y') = \|\psi(y) - \psi(y')\|^2_{H_{\mathcal{Y}}}$ for an embedding $\varphi : \mathcal{X} \to H_{\mathcal{X}}$ in a Hilbert space $H_{\mathcal{X}}$ and similarly for $\mathcal{Y}$, computation (3) holds. Such a property depends on the metric space, and when it is satisfied, the metric space is said to be of *negative type* or the distance to be *Hilbertian*. Another equivalent formulation is to say that $d_{\mathcal{X}}$ is a *conditionally negative definite kernel* on $\mathcal{X}$. We refer to [Lyo13] for a thorough discussion.

Definition 3. *A function $k_{\mathcal{X}} : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ is a* conditionally negative definite (CND) kernel *if it is symmetric and for all $N \geq 1$, $x_1, \dots, x_N \in \mathcal{X}$ and $\omega_1, \dots, \omega_N \in \mathbb{R}$ such that $\sum_{i=1}^N \omega_i = 0$, $\sum_{i,j\leq N} \omega_i\omega_j k_{\mathcal{X}}(x_i, x_j) \leq 0$.*

Every CND kernel can be written as $k_{\mathcal{X}}(x,x') = f(x) + f(x') + \frac{1}{2}\|\varphi(x) - \varphi(x')\|^2_H$ for an embedding $\varphi : \mathcal{X} \to H$ with H a Hilbert space and a function f, as shown in [Sch38]. As far as the Gromov–Wasserstein functional is concerned, our discussion above shows that $c_{\mathcal{X}}$ can actually be replaced with a kernel which is CND and that the relaxation still holds. To sum up our review of the literature,

Proposition 1. Let $(\mathcal{X}, k_{\mathcal{X}}, \mu)$ and $(\mathcal{Y}, k_{\mathcal{Y}}, \nu)$ be two spaces endowed each with a conditionally negative definite kernel (or each with a conditionally positive definite one) and a probability measure. Then the bi-convex relaxation of GW_2^2 is tight, in the sense of (2). The corresponding kernel $k((x,y),(x',y'))$ gives a non-positive quadratic form on signed measures with null marginals on $\mathcal{X}\times\mathcal{Y}$.

Remark that the map $(\pi,\gamma) \mapsto F(\pi,\gamma)$ is indeed bi-convex as it is linear in each variable π, γ. Obviously, due to the square in the quadratic GW problem, the "sign" of both kernels do not matter. There are several important Riemannian manifolds which are of negative type; among them the real Hyperbolic space, the sphere and the Euclidean space. Counter-examples are for instance in finite dimension the Hyperbolic space on the quaternions [FH74], and in infinite dimension the Wasserstein-2 distance in $\mathbb{R}^d$ for $d \geq 3$ as proved in [ANN18].

1.2 Related work

1.2.1 Monge maps for the OT problem. The (OT) problem has been extensively studied (see [San15, Vil08, PC19] for a thorough introduction) and particular attention has been devoted to situations where existence of Monge maps, or variations of, can be ensured.

Brenier's theorem [Bre87], stated below, is the most well-known of such cases where the optimal plan is a map.

Theorem 1 (Brenier's theorem)**.** Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ such that the optimal cost between μ and ν is finite and $c(x,y) = |x-y|^2$. If $\mu \ll \mathcal{L}^d$, then there exists a unique (up to a set of μ-measure zero) solution of (OT) and it is induced by a map T. This map is characterized by being the unique gradient of a convex function $T = \nabla f$ such that $(\nabla f)_{\#}\mu = \nu$.

This central result admits a generalization in the manifold setting, initially proposed by [McC01], that we shall use later on:

Proposition 2 ([Vil08, Thm. 10.41])**.** Let M be a Riemannian manifold with distance function d, and $c(x,y) = d(x,y)^2$. Let $\mu, \nu \in \mathcal{P}(M)$ with compact support. If $\mu \ll \mathrm{vol}_M$, then there exists a unique solution of (OT) and it can be written as

$$y = T(x) = \exp_x(\nabla f(x)),$$

where f is some $d^2/2$-convex function.

Brenier's theorem can be extended in a few directions. The condition that μ has a density can be weakened to the fact that it does not give mass to sets of Hausdorff dimension smaller than $d-1$ (*e.g.* hypersurfaces), and c can be more general than the squared distance function, as long as it satisfies the *twist condition*, defined below. In the following, $\mathcal{X} = \mathcal{Y}$ are complete Riemannian manifolds and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a continuous cost function, differentiable w.r.t. x. We refer to [MG11, Vil08] for more information on the twist condition, to [CMN10, AKM11, McC12] for the introduction of the subtwist condition and to [Moa16] for that of the m-twist and generalized twist conditions.

Proposition 3 (Twist)**.** We say that c satisfies the *twist condition* if

$$\text{for all } x_0 \in \mathcal{X}, \quad y \mapsto \nabla_x c(x_0, y) \in T_{x_0}\mathcal{X} \text{ is injective.} \tag{Twist}$$

Suppose that c satisfies (Twist) and assume that any c-concave function is differentiable μ-a.e. on its domain. If μ and ν have finite transport cost, then (OT) admits a unique optimal transport plan $\pi^\star$ and it is induced by a map which is the gradient of a c-convex function $f : \mathcal{X} \to \mathbb{R}$:

$$\pi^\star = (\mathrm{id}, c\text{-}\exp_x(\nabla f))_\# \mu \,.$$

Remark 1. Following [MG11, Vil08], we recall that the *c-exponential map* is defined on the image of $-\nabla_x c$ by the formula $c\text{-}\exp_x(p) = (\nabla_x c)^{-1}(x, -p)$, *i.e.* $c\text{-}\exp_x(p)$ is the unique y such that $\nabla_x c(x, y) + p = 0$. This notion reduces to the usual Riemannian exponential map when $c = d^2/2$.

Remark 2. Costs of the form $c(x, y) = h(x - y)$ with h strictly convex, and in particular the costs $c(x, y) = |x - y|^p$ for $p > 1$, do satisfy the twist condition. Unfortunately, it cannot be satisfied for smooth costs on compact manifolds.

The twist condition is equivalent to the fact that for all $y_1 \neq y_2 \in \mathcal{Y}$, the function $x \in \mathcal{X} \mapsto c(x, y_1) - c(x, y_2)$ has no critical point. Two weaker notions can be introduced in order to retain some (but less) structure on the optimal plans:

Proposition 4 (Subtwist)**.** We say that c satisfies the *subtwist condition* if

$$\begin{aligned}&\text{for all } y_1 \neq y_2 \in \mathcal{Y}, x \in \mathcal{X} \mapsto c(x, y_1) - c(x, y_2) \text{ has no critical points,}\\ &\text{save at most one global maximum and at most one global minimum.}\end{aligned} \tag{Subtwist}$$

Suppose that c satisfies (Subtwist) and is bounded. If μ vanishes on each hypersurface, then (OT) admits a unique optimal transport plan $\pi^\star$ and it is induced by the union of a map and an anti-map:

$$\pi^\star = (\mathrm{id}, G)_\# \bar{\mu} + (H, \mathrm{id})_\#(\nu - G_\# \bar{\mu})$$

for some (Borel) measurable maps $G : \mathcal{X} \to \mathcal{Y}$ and $H : \mathcal{Y} \to \mathcal{X}$ and non-negative measure $\bar{\mu} \leq \mu$ such that $\nu - G_\# \bar{\mu}$ vanishes on the range of G.

Proposition 5 (m-twist)**.** We say that c satisfies a *m-twist* (resp. *generalized twist*) *condition* if

$$\text{for all } (x_0, y_0) \in \mathcal{X} \times \mathcal{Y}, \text{ the set } \{y \mid \nabla_x c(x_0, y) = \nabla_x c(x_0, y_0)\} \text{ has at most } m \text{ elements} \tag{m-twist}$$

(resp. is a finite subset of $\mathcal{Y}$). Suppose that c is bounded, satisfies (m-twist) and assume that any c-concave function is differentiable μ-almost surely on its domain. If μ has not atom and μ and ν have finite transport cost, then each optimal plan $\pi^\star$ of (OT) is supported on the graphs of $k \in [\![m]\!]$ (resp. in $\mathbb{N} \cup \{\infty\}$) measurable maps, *i.e.* there exists non-negative functions $\alpha_i : \mathcal{X} \to [0, 1]$ and (Borel) measurable maps $T_i : \mathcal{X} \to \mathcal{Y}$ such that

$$\pi^\star = \sum_{i=1}^{k} \alpha_i (\mathrm{id}, T_i)_\# \mu \,,$$

in the sense $\pi^\star(S) = \sum_{i=1}^{k} \int_{\mathcal{X}} \alpha_i(x) \mathbf{1}_S(x, T_i(x)) \, \mathrm{d}\mu(x)$ for any Borel $S \subset \mathcal{X} \times \mathcal{Y}$.

Example 1. If $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and $c(x, y) = x^2y^2 + \lambda xy$ for some $\lambda \neq 0$, the 2-twist condition holds. As we shall see in Section 3.3, such costs are closely related to the quadratic GW problem (GW-Q) in dimension 1.

Remark 3. Following [Vil08, Rem. 10.33], when measures μ and ν have compact support and μ has a density—which are assumptions that we make in the following—, all conditions of Propositions 3 to 5 are satisfied.

Additionally, when $\mathcal{X}$ and $\mathcal{Y}$ are of dimension d and μ and ν are compactly supported, [MPW12] shows that if c is C^2 and satisfies the weaker condition of *non-degeneracy* $\det D^2_{xy}c(x, y) \neq 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, any solution of (OT) is supported on a d-dimensional Lipschitz submanifold; but with no guarantee of uniqueness. General conditions for the uniqueness of the (OT) problem to hold have been recently established in [MR20].

1.2.2 Monge maps for the GW problem. Besides being an interesting mathematical question in itself and in addition to reducing the computational complexity of the problem, preferring Monge maps in the Gromov–Wasserstein setting is interesting for registration purposes via mappings in imaging, shape analysis and simulation based inference; see [HGL21]. The interest of restricting the Gromov–Wasserstein problem to the class of mappings is discussed in [MN22], where it is shown that this formulation retains some properties of a metric [MN22, Thm. 3].

In sharp contrast with the optimal transportation problem, there are very few results that ensure the existence of a Monge map for the Gromov–Wasserstein problem, even in the particular cases considered in this work. For the inner product cost, [Vay20, Thm. 4.2.3] provides a positive answer for the scalar product cost under some assumptions which are nontrivial to check.

Proposition 6 (Inner product cost: optimal map under condition)**.** Let $n \geq d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures of finite second order moment with $\mu \ll \mathcal{L}^n$. Suppose that there exists $\pi^\star$ solution of (GW-IP) such that $M^\star \triangleq \int y \otimes x \, \mathrm{d}\pi^\star(x, y)$ is of full rank. Then there exists an optimal map between μ and ν that can be written as $T = \nabla f \circ M^\star$ with $f : \mathbb{R}^d \to \mathbb{R}$ convex.

For the quadratic cost, in [Vay20] is claimed that in the discrete case in dimension 1 with uniform mass and same number of points N, the optimal solution of (QAP) would either be the identity $\sigma(i) = i$ or the anti-identity $\sigma(i) = N + 1 - i$ (Thm. 4.1.1). However, a counter-example to this assertion has been recently provided by [BHS22]. To the best of our knowledge, the only positive results on the existence of Monge maps for the quadratic cost are the following.

Proposition 7 ([Stu12, Thm. 9.21])**.** Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$. Assume that $\mu, \nu \ll \mathcal{L}^n$ and that both measures are rotationally invariant around their barycenter. Then every $\pi^\star$ solution of (GW-Q) is induced by a transport map T, unique up to composition with rotations, which rearranges the radial distributions of μ and ν around their barycenter in a monotone non-decreasing way.

It is of interest to note that this result implies that the optimal transport map between two Gaussians in dimension one is a monotone map. Yet, in higher dimension, the optimal plan for Gaussians is not known unless more constraints on the structure of the optimal plan is imposed [SDD21].

Proposition 8 ([Vay20, Prop. 4.2.4])**.** Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ with compact support, with $n \geq d$. Assume that $\mu \ll \mathcal{L}^n$ and that both μ and ν are centered. Suppose that there exists $\pi^\star$ solution of (GW-Q) such that $M^\star = \int y \otimes x \, \mathrm{d}\pi^\star(x, y)$ is of full rank. Then there exists $f : \mathbb{R}^d \to \mathbb{R}$ convex such that $T = \nabla f \circ M^\star$ pushes μ to ν. Moreover, if there exists a differentiable convex $F : \mathbb{R} \to \mathbb{R}$ such that $|T(x)|_2^2 = F'(|x|_2^2)$ μ-a.e., then T is optimal for (GW-Q).

These results ensure the existence of a Monge map for the GW problems under restrictive conditions, either on the symmetry of the measures μ and ν or on the rank of a matrix that depends on an optimal correspondence plan.

1.3 Outline, contributions and perspectives

This work is organized in the following way. Section 2 provides a general setting in which existence of optimal *transport* maps can be shown for optimal transport problems with costs that are defined by submersions. We provide two versions of the result, one (Theorem 2) which has no structure and is fairly general and one (Theorem 3) which imposes a more structured setting thus recovering more structure in the optimal maps; the latter having the benefit of being more usable in further theoretical or computational analysis. The proof of the second version requires a measurability argument which is addressed in details in Proposition 11. Applications of these results to the Gromov–Wasserstein problems in Euclidean spaces with cost being the inner product (GW-IP) or the squared Euclidean distance (GW-Q) are done in Section 3. We prove existence of Monge maps for the scalar product, improving on the result in [Vay20] by alleviating the need for the rank assumption in Proposition 6, difficult to check in practice. For the squared distance, the worst case, but probably the generic one, is a bi-map structure, yet Monge maps can occur as well in a characterized setting.

An important question is to know if the result in the quadratic case (GW-Q) is sharp, that is we cannot guarantee that a Monge map exists in general, but only a bi-map. We address it in Section 3.3 which focuses on the one-dimensional case with quadratic cost and consists into two parts: first, after proving some structural properties of the optimal plan, we conduct a numerical exploration that suggests our previous structural results are sharp in dimension one. Last, we prove a positive result on the optimality of monotone maps, which partly explains why such maps are often optimal in practice. Notably, this result highlights the importance of long-range effects of this cost. In comparison, Sturm's result (Proposition 7) is concerned with symmetric distributions and does not account for the phenomenon that very often in dimension 1, a monotone map is optimal. As a possible extension, it is of interest to know if this result is stable with respect to the input distributions. We show that it is not the case in dimension 1, *i.e.* that there exists (degenerate) symmetric cases (namely, $\mu = \nu = \delta_0$) for which a monotone map is not optimal for perturbations of the measures (*i.e.* there exist $\mu_n, \nu_n \to \mu, \nu$ for which the optimal correspondence plan is never a monotone map)—although this result does not prevent stability in non-degenerate symmetric cases. Further extensions could address sharper sufficient conditions for guaranteeing optimality of monotone plans.

Left open by our work is the existence of a cost which gives a proper Gromov–Wasserstein distance, contrary to the scalar product, and which guarantees the existence of Monge maps. However, for practical applications, having the possibility to choose the cost regardless of whether it defines a distance, for instance to favor local or long range effects, may be highly relevant since it directly affects the optimal correspondence plans, as well as the optimization problem.

2 Existence of Monge maps for fiber-invariant costs

This section provides the main result on the existence of Monge maps for OT problems for which the cost satisfies an invariance property. As detailed in Section 3, this property will be satisfied by the transport costs $C_{\pi^\star}$ arising from the first-order condition of (GW-Q) and (GW-IP)—see Section 1.1.2.

2.1 Statement of the results

We briefly present the main idea. Let μ, ν be two probability measures supported on a measurable space (E, Σ_E) and consider a measurable map $\varphi : E \to B$, for some measurable space (B, Σ_B), referred to as the *base space* latter on. We shall omit to mention the σ-algebra afterwards. Let $(\mu_u)_{u\in B}$ (resp. $(\nu_u)_{u\in B}$) denote a disintegration of μ (resp. ν) with respect to φ (see Appendix A.3 or for instance [AGS05, Thm. 5.3.1]). Let $c : E \times E \to \mathbb{R}$ that is invariant on the fibers of φ, that is $c(x, y) = \tilde{c}(\varphi(x), \varphi(y))$ for all $(x, y) \in E \times E$ and some cost function $\tilde{c}$ on $B \times B$. Solving the OT problem between μ and ν for c reduces to the OT problem between $\varphi_\# \mu$ and $\varphi_\# \nu$ on $B \times B$ for $\tilde{c}$. Assuming that there exists a Monge map t_B between $\varphi_\# \mu$ and $\varphi_\# \nu$, we build a Monge map T between μ and ν by (i) transporting each fiber μ_u onto $\nu_{t_B(u)}$ using *a map* T_u, and (ii) gluing the $(T_u)_{u\in B}$ together to define a *measurable* map T

satisfying $T_\#\mu = \nu$ that will be optimal as it coincides with t_B on B and the cost c does not depend on the fibers $(\varphi^{-1}(u))_{u\in B}$. We underline that ensuring the measurability of the map T is non-trivial and crucial from a theoretical standpoint.

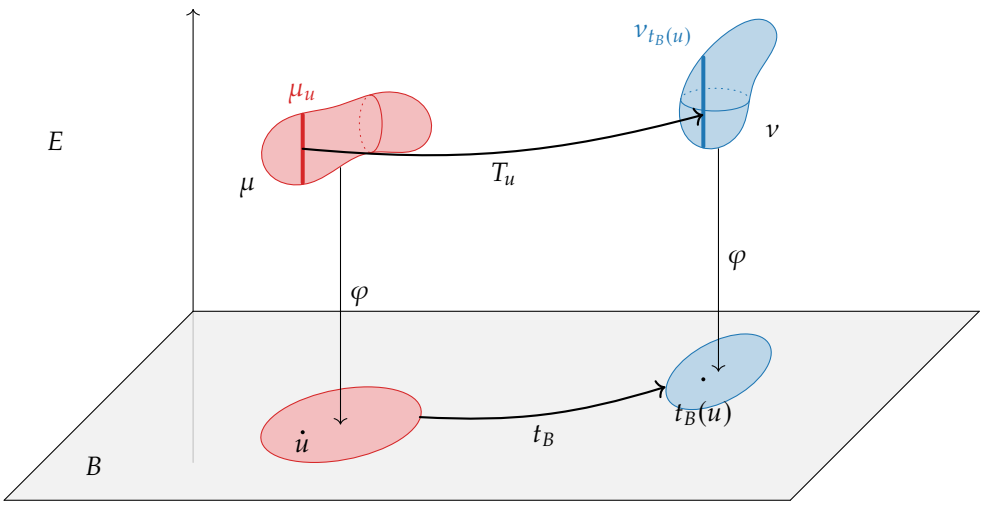


Figure 1: Illustration of the construction of a Monge map between μ and ν: we optimally transport the projections of the measures in B and then "lift" the resulting map t_B to E by sending each fiber μ_u onto the fiber $\nu_{t_B(u)}$, resulting respectively from the disintegrations of μ and ν by φ.

We propose two theorems to formalize this idea: the first one guarantees in a fairly general setting the existence of a Monge map for the (GW) problem, but its construction is quite convoluted and there is little to no hope that it can be leveraged in practice, either from a theoretical or computational perspective. Assuming more structure, in particular on the fibers of φ, enables the construction of a Monge map for (GW) with a structure akin to Proposition 2. As detailed in Section 3, both (GW-Q) and (GW-IP) fall in the latter setting.

Theorem 2. Let $\mathcal{X}$ and $\mathcal{Y}$ be two measurable spaces for which there exist two measurable maps $\Phi_\mathcal{X} : \mathcal{X} \to \mathbb{R}^d$ and $\Phi_\mathcal{Y} : \mathcal{Y} \to \mathbb{R}^d$ that are injective, and whose inverses are measurable. Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be two probability measures. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a cost function, and B_+, B_- be two measurable spaces along with measurable maps $\varphi : \mathcal{X} \to B_+$ and $\psi : \mathcal{Y} \to B_-$. Assume that there exists a cost $\tilde{c} : B_+ \times B_- \to \mathbb{R}$ such that

$$c(x, y) = \tilde{c}(\varphi(x), \psi(y)) \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}$$

and that there exists a Monge map $t_B : B_+ \to B_-$ that transports $\varphi_\#\mu$ onto $\psi_\#\nu$ for the cost $\tilde{c}$. Assume that there exists a disintegration $(\mu_u)_{u\in B_+}$ of μ with respect to φ such that $\varphi_\#\mu$-a.e., μ_u is atomless. Then there exists a Monge map between μ and ν for the cost c. Furthermore, it projects onto t_B through (φ, ψ), in the sense that $(\varphi, \psi)_\#(\mathrm{id}, T)_\#\mu = (\mathrm{id}, t_B)_\#(\varphi_\#\mu)$.

The proof of this theorem is provided in Section 2.2.

Remark 4. The atomless assumption on the disintegration $(\mu_u)_u$ is a natural minimal requirement to expect the existence of a map (without specific assumption on the target measure ν) and implies in particular that the fibers $(\varphi^{-1}(u))_{u\in B_+}$ should not be discrete (at least $\varphi_\#\mu$-a.e.). Indeed, if for instance $\mathcal{X} = \mathcal{Y} = B_+ = B_- = \mathbb{R}$ and $\varphi : x \mapsto |x|$, the fibers of φ are of the form $\{-u, u\}$, hence the disintegrations $(\mu_u)_{u\geq 0}$ and $(\nu_u)_{u\geq 0}$ are discrete and given by $\mu_u(u)\delta_u + (1-\mu_u(u))\delta_{-u}$ and $\nu_u(u)\delta_u + (1-\nu_u(u))\delta_{-u}$, and there is in general no map T_u between two such discrete measures, unless we assume that $\mu_u(u) = \nu_u(u)$ or $1 - \nu_u(u)$, $\varphi_\#\mu$-a.e.

Observe also that $\varphi_\#\mu$ may have atoms: as we assume the existence of the Monge map t_B, it would imply in that case that $\psi_\#\nu$ also has atoms.

Remark 5. The "projection" property $(\varphi,\psi)_\#(\mathrm{id},T)_\#\mu = (\mathrm{id},t_B)_\#(\varphi_\#\mu)$ can also be written $\psi\circ T(x) = t_B\circ\varphi(x)$, for μ-a.e. x. A converse implication, that is "every Monge map between μ and ν projects onto a Monge map between $\varphi_\#\mu$ and $\varphi_\#\nu$" may not hold in general. This is however true if we can guarantee that there is a unique optimal transport plan between $\varphi_\#\mu$ and $\psi_\#\nu$ and that it is of the form $(\mathrm{id},t_B)_\#\mu$ (*e.g.* if we can apply Theorem 1)—in that case, T necessary projects onto t_B in the aforementioned sense.

Under additional assumptions, we can build a more structured Monge map. Namely, as our goal is to apply Proposition 2, we will assume that the (common) basis $B = B_+ = B_-$ is a manifold, that *almost all* the fibers of $\varphi : E \to B$ are homeomorphic to the *same* manifold F, and that every source measure of interest $(\mu, \mu_u, \varphi_\#\mu)$ has a density. We also introduce the following convention: if $\mu\in\mathcal{P}(E)$ for some measurable space E, $E'\subset E$, and $\varphi : E'\to B$, we let $\varphi_\#\mu$ be the (non-negative) measure supported on B defined by $\varphi_\#\mu(A) = \mu(\varphi^{-1}(A))$ for $A\subset B$ measurable. If $\mu(E') = 1$, note that $\varphi_\#\mu$ defines a probability measure on B. This formalism allows us to state our theorem even when some assumptions only hold μ-a.e.

Theorem 3. Let E_0 be a measurable space and B_0 and F be complete Riemannian manifolds. Let $\mu,\nu\in\mathcal{P}(E_0)$ be two probability measures with compact support. Assume that there exists a set $E\subset E_0$ such that $\mu(E) = 1$ and that there exists a measurable map $\Phi : E\to B_0\times F$ that is injective and whose inverse on its image is measurable as well. Let p_B, p_F denote the projections of $B_0\times F$ on B_0 and F respectively, and let $\varphi \triangleq p_B\circ\Phi : E\to B_0$. Let $c : E_0\times E_0\to\mathbb{R}$ and suppose that there exists a twisted $\tilde{c} : B_0\times B_0\to\mathbb{R}$ such that

$$c(x,y) = \tilde{c}(\varphi(x),\varphi(y)) \quad \text{for all } (x,y)\in E_0\times E_0\,.$$

Assume that $\varphi_\#\mu$ is absolutely continuous w.r.t. the Lebesgue measure on B_0 and let thus t_B denote the unique Monge map between $\varphi_\#\mu$ and $\varphi_\#\nu$ for this cost. To alleviate notations, we let $\mu'\triangleq\Phi_\#\mu$ and $\nu'\triangleq\Phi_\#\nu$ in the following. Suppose that there exists a disintegration $(\mu'_u)_{u\in B_0}$ of μ' by p_B such that for $\varphi_\#\mu$-a.e. u, μ'_u is absolutely continuous w.r.t. the volume measure on F.
Then there exists an optimal map T between μ and ν for the cost c that can be decomposed as

$$\Phi\circ T\circ\Phi^{-1}(u,v) = (t_B(u), t_F(u,v)) = (\tilde{c}\text{-}\exp_u(\nabla f(u)), \exp_v(\nabla g_u(v)))\,, \tag{3}$$

with $f : B_0\to\mathbb{R}$ $\tilde{c}$-convex and $g_u : F\to\mathbb{R}$ $d_F^2/2$-convex for $\varphi_\#\mu$-a.e. u. Note that t_F could actually be any measurable function that sends each fiber μ'_u onto $\nu'_{t_B(u)}$.

The proof of this theorem is provided in Section 2.3. Let us give a simple example that illustrates the role played by our assumptions. This example has connections with (GW-Q) as detailed in Section 3.2.

Example 2. Let $E_0 = \mathbb{R}^d$ and $E = E_0\backslash\{0\}$, let $B_0 = \mathbb{R}$ and $F = S^{d-1} = \{x\in E_0 \mid |x| = 1\}$. For convenience, we also introduce the space $B = \mathbb{R}_{>0}$. Consider the cost function $c(x,y) = (|x|-|y|)^2$, so that c only depends on the norm of its entries. The fibers of the map $x\mapsto|x|$ are spheres, with the exception of $x = 0$, which invites us to consider the diffeomorphism

$$\begin{aligned}\Phi : E &\to \mathbb{R}_{>0}\times S^{d-1} = B\times F\subset B_0\times F\\ x &\mapsto \left(|x|, \frac{x}{|x|}\right).\end{aligned}$$

From this, we can write $c(x,y) = \tilde{c}(\varphi(x),\varphi(y))$ where $\varphi(x) = |x|$ and $\tilde{c}(u,u') = (u-u')^2$ (which is twisted).

Now, if μ has a density on $\mathbb{R}^d$, so does μ' on $B_0\times F$ as Φ is a diffeomorphism. The coarea formula gives the existence of a disintegration $(\mu'_u)_{u\in B}$ of μ' by $p_B : (u,v)\mapsto u$ (note that $p_{B\#}\mu' = \varphi_\#\mu$ also has a density) such that all the μ'_u admit a density on S^{d-1}.

Our theorem thus applies, ensuring the existence of a structured Monge map between μ and (any) ν for the cost c: it decomposes for almost all $x \in \mathbb{R}^d$ as a Monge map on the basis $B_0 = \mathbb{R}$ (although it is actually only characterized on the image of φ, that is $B = \mathbb{R}_{>0}$) obtained as the gradient of a convex function f (there is no need for the exponential map here and ∇f is the non-decreasing mapping between the quantiles of $\varphi_\# \mu$ and $\varphi_\# \nu$) and a Monge map on each fiber $F = S^{d-1}$, also built from gradients of convex functions (via the exponential map on the sphere).

Note that our theorem only requires assumptions to hold almost everywhere on $E_0 = \mathbb{R}^d$, which is important since it allows to ignore the singularity of φ at $x = 0$.

2.2 Proof of Theorem 2

The proof decomposes in three steps.

Step 1: Existence and optimality of lifts. We know by assumption that there exists a Monge map t_B that is optimal between the pushforward measures $\varphi_\# \mu$ and $\psi_\# \nu$.

As our goal is to build a Monge map between the initial measures μ and ν, we first show that (i) there exists a transport plan $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_\# \pi = (\mathrm{id}, t_B)_\# \mu$ and that (ii) any such π is an optimal transport plan between μ and ν for the cost c. This is formalized by the following lemmas.

Lemma 1 (Existence of a lift)**.** For any transport plan $\tilde{\pi} \in \Pi(\varphi_\# \mu, \psi_\# \nu)$, there exists a transport plan $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_\# \pi = \tilde{\pi}$.

Proof. Let $(\mu_u)_{u \in B_+}$ and $(\nu_v)_{v \in B_-}$ be disintegrations of μ and ν by φ and ψ respectively. Given $\tilde{\pi} \in \Pi(\varphi_\# \mu, \psi_\# \nu)$, we define

$$\pi \triangleq \iint_{B_+ \times B_-} (\mu_u \otimes \nu_v) \, \mathrm{d}\tilde{\pi}(u, v) \,,$$

which is in $\Pi(\mu, \nu)$ since $\tilde{\pi}$ has $(\varphi_\# \mu, \psi_\# \nu)$ as marginals and, by disintegration, $\int_{B_+} \mu_u \, \mathrm{d}(\varphi_\# \mu) = \mu$ (and same for ν). The relation $(\varphi, \psi)_\# \mu = \tilde{\pi}$ then follows from the fact that by disintegration μ_u is exactly supported on $\varphi^{-1}(\{u\})$ so one has $\mu_u(\varphi^{-1}(U)) = \delta_U(u)$ for any measurable $U \subset B_+$ (and same for ν_v). □

Lemma 2. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $\tilde{c} : B_+ \times B_- \to \mathbb{R}$ such that

$$c(x, y) = \tilde{c}(\varphi(x), \psi(y)) \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Then optimal plans for the base space cost $\tilde{c}$ are the projections of optimal plans for c: writing $\Pi_c^\star(\mu, \nu)$ the set of optimal transport plan between μ and ν for the cost c, and similarly for $\Pi_{\tilde{c}}^\star(\varphi_\# \mu, \psi_\# \nu)$, one has

$$\Pi_{\tilde{c}}^\star(\varphi_\# \mu, \psi_\# \nu) = (\varphi, \psi)_\# \Pi_c^\star(\mu, \nu) \,.$$

Proof. Let us first remark that the relation between c and $\tilde{c}$ implies that for any $\pi \in \Pi(\mu, \nu)$ and $\tilde{\pi} = (\varphi, \psi)_\# \pi$, one has $\langle c, \pi \rangle = \langle \tilde{c}, \tilde{\pi} \rangle$. Now, let $\tilde{\pi}^\star \in \Pi_{\tilde{c}}^\star(\varphi_\# \mu, \psi_\# \nu)$; by Lemma 1, there exists a $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_\# \pi = \tilde{\pi}^\star$. Then for any $\gamma \in \Pi(\mu, \nu)$, one has $\langle c, \pi \rangle = \langle \tilde{c}, \tilde{\pi}^\star \rangle \leq \langle \tilde{c}, (\varphi, \psi)_\# \gamma \rangle = \langle c, \gamma \rangle$, hence the optimality of π. Conversely, let $\pi^\star \in \Pi_c^\star(\mu, \nu)$; by Lemma 1, for any $\tilde{\gamma} \in \Pi(\varphi_\# \mu, \psi_\# \nu)$ there exists a $\gamma \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_\# \gamma = \tilde{\gamma}$. We then have $\langle \tilde{c}, (\varphi, \psi)_\# \pi^\star \rangle = \langle c, \pi^\star \rangle \leq \langle c, \gamma \rangle = \langle \tilde{c}, \tilde{\gamma} \rangle$, hence the optimality of $(\varphi, \psi)_\# \pi^\star$. □

Step 2: Existence of Monge maps between the fibers. Using Lemma 1 with $\tilde{\pi} = (\mathrm{id}, t_B)_\#(\varphi_\# \mu)$, we know that we can build an optimal transportation plan $\pi \in \Pi(\mu, \nu)$ that essentially coincides with t_B on $B_+ \times B_-$ and transports each fiber μ_u onto $\nu_{t_B(u)}$ for μ-a.e. $u \in B_+$. In order to build a Monge map between μ and ν, we must show that one can actually transport almost all μ_u onto $\nu_{t_B(u)}$ using a map rather than a plan. For this, we use the following result, see [San15, Rem. 1.23, Lemma 1.28, Cor. 1.29].

Proposition 9. Let α, β be two measures supported on $\mathbb{R}^d$ with α atomless. Then:
(i) if $d = 1$, there exists a transport map $\tilde{T}$ that pushes α onto β. Furthermore, it is the *unique* optimal map between these measures for the quadratic cost $(x, y) \mapsto |x - y|^2$;
(ii) there exists a map $\sigma_d : \mathbb{R}^d \to \mathbb{R}$ (that does not depend on α, β) that is (Borel) measurable, injective, and its inverse is measurable as well.

As we assumed that the ground spaces $\mathcal{X}$ and $\mathcal{Y}$ can be embedded in $\mathbb{R}^d$ using the injective, measurable maps $\Phi_{\mathcal{X}}$ and $\Phi_{\mathcal{Y}}$, we can apply Proposition 9 using $\sigma_{\mathcal{X}} = \sigma_d \circ \Phi_{\mathcal{X}}$ and $\sigma_{\mathcal{Y}} = \sigma_d \circ \Phi_{\mathcal{Y}}$. As $\sigma_{\mathcal{X}}$ is injective, $\sigma_{\mathcal{X}\#}\mu_u$ is atomless on $\mathbb{R}$, and we can thus consider the *unique* Monge map $\tilde{T}_u$ between $\sigma_{\mathcal{X}\#}\mu_u$ and $\sigma_{\mathcal{Y}\#}\nu_{t_B(u)}$ for the quadratic cost on $\mathbb{R}$.

From this, as the maps $\sigma_{\mathcal{X}}$ and $\sigma_{\mathcal{Y}}$ are measurable and injective (thus invertible on their image) we can define $T_u = \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}_u \circ \sigma_{\mathcal{X}} : \mathcal{X} \to \mathcal{Y}$, that defines a (measurable) transport map between μ_u and $\nu_{t_B(u)}$.

Step 3: building a measurable global map. Now that we have maps $(T_u)_u$ between each μ_u and $\nu_{t_B(u)}$, it may be tempting to simply define a map $T : \mathcal{X} \to \mathcal{Y}$ by $T(x) = T_{\varphi(x)}(x)$ when $\mu_{\varphi(x)}$ is atomless (which, by assumption, holds μ-a.e.). Intuitively, this map induces a transport plan $(\mathrm{id}, T)_\#\mu$ that satisfies $(\varphi, \psi)_\#(\mathrm{id}, T)_\#\mu = (\mathrm{id}, t_B)_\#(\varphi_\#\mu)$ on $B_+ \times B_-$ and thus must be optimal according to Lemma 2.

One remaining step, though, is to prove that this map T can be defined in a measurable way. For this, we use the following *measurable selection theorem* due to [FGM10, Thm. 1.1], that reads:

Proposition 10. Let (B, Σ, m) be a σ-finite measure space and consider a measurable function $B \ni u \mapsto (\mu_u, \nu_u) \in \mathcal{P}(\mathbb{R}^d)^2$. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function, and assume that for m-a.e. $u \in B$, there is a (unique) Monge map T_u between μ_u and ν_u for the cost c.
Then there exists a measurable function $(u, x) \mapsto T(u, x)$ such that m-a.e., $T(u, x) = T_u(x)$, μ_u-a.e.

We can apply this result in the case $d = 1$ to the family of measures $(\sigma_{\mathcal{X}\#}\mu_u, \sigma_{\mathcal{Y}\#}\nu_{t_B(u)})_{u \in B_+}$, where the reference measure on B_+ is $\varphi_\#\mu$.[1] We first need to show the measurability of this family of measures. By definition of the disintegration of measures, the map $v \in B_- \mapsto \nu_v$ is measurable; and as the Monge map t_B is measurable as well, so is the map $B_+ \ni u \mapsto \sigma_{\mathcal{Y}\#}v_{t_B(u)}$ by composition of measurable maps, and thus so is the map $u \mapsto (\mu_u, \nu_{t_B(u)})$. Proposition 10 therefore applies and guarantees the existence of a measurable map $\tilde{T} : B_+ \times \mathbb{R} \to \mathbb{R}$ such that $\tilde{T}(u, x) = \tilde{T}_u(x)$ for $\varphi_\#\mu$ almost all u and $\sigma_{\mathcal{X}\#}\mu$ almost all x. Now, we can define

$$\begin{aligned} T : \mathcal{X} &\to \mathcal{Y} \\ x &\mapsto \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}(\varphi(x), \sigma_{\mathcal{X}}(x)) \,. \end{aligned}$$

This map is measurable as composition of measurable maps. Let us prove that this defines a transport map between μ and ν. For any function $\zeta : \mathcal{Y} \to \mathbb{R}$ continuous with compact support, we can write

$$\int_{\mathcal{Y}} \zeta(y)\, \mathrm{d}T_\#\mu(y) = \int_{\mathcal{X}} \zeta(T(x))\, \mathrm{d}\mu(x) = \int_{u \in B_+} \int_{x \in \varphi^{-1}(\{u\})} \zeta\left(\sigma_{\mathcal{Y}}^{-1}\left(\tilde{T}_u(\sigma_{\mathcal{X}}(x))\right)\right) \mathrm{d}\mu_u(x)\, \mathrm{d}\varphi_\#\mu(u) \,,$$

where we use the disintegration of μ w.r.t. φ and the fact that the μ_u are supported on $\varphi^{-1}(\{u\})$, allowing us to write $\tilde{T}(\varphi(x), \sigma_{\mathcal{X}}(x)) = \tilde{T}_u(\sigma_{\mathcal{X}}(x))$ on that fiber ($\varphi_\#\mu$-a.e.). Now, recall that $T_u : x \mapsto \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}_u \circ \sigma_{\mathcal{X}}(x)$ defines a transport map between μ_u and $\nu_{t_B(u)}$. In particular, the image of the fiber $\varphi^{-1}(\{u\})$ by this map is $\psi^{-1}(\{t_B(u)\}) \subset \mathcal{Y}$. Since $\nu_{t_B(u)}$ is supported on $\psi^{-1}(\{t_B(u)\})$, we therefore get

$$\int_{\mathcal{Y}} \zeta(y)\, \mathrm{d}T_\#\mu(y) = \int_{u \in B_+} \int_{y \in \mathcal{Y}} \zeta(y)\, \mathrm{d}\nu_{t_B(u)}(y)\, \mathrm{d}\varphi_\#\mu(u) = \int_{v \in B_-} \int_{y \in \mathcal{Y}} \zeta(y)\, \mathrm{d}\nu_v(y)\, \mathrm{d}\psi_\#\nu(v) \,,$$

[1] Note that we cannot apply Proposition 10 to the measures $(\mu_u, \nu_{t_B(u)})_u$ and the maps $(T_u)_u$ directly, as T_u may not be the unique Monge map between the measures, a required assumption of the proposition.

where we use the fact that t_B pushes $\varphi_\#\mu$ to $\psi_\#\nu$; by disintegration of ν w.r.t. ψ, we then have $T_\#\mu = \nu$.

By Lemma 2, this map is optimal if and only if it satisfies $(\varphi,\psi)_\#(\mathrm{id},T)_\#\mu = (\mathrm{id},t_B)_\#(\varphi_\#\mu)$, as t_B is an optimal transportation plan between $\varphi_\#\mu$ and $\psi_\#\nu$, making $(\mathrm{id},T)_\#\mu$ optimal between μ and ν (hence T a Monge map). For this, let $\zeta : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ be a continuous function with compact support. We have

$$\begin{aligned}
\int_{B_+\times B_-} \zeta(u,v)\,\mathrm{d}(\varphi,\psi)_\#(\mathrm{id},T)_\#\mu(u,v) &= \int_{\mathcal{X}} \zeta(\varphi(x),\psi(T(x)))\,\mathrm{d}\mu(x)\\
&= \int_{u\in B_+}\int_{x\in\varphi^{-1}(\{u\})} \zeta\left(u,\psi(\sigma_{\mathcal{Y}}^{-1}\circ\tilde{T}_u\circ\sigma_{\mathcal{X}}(x))\right)\,\mathrm{d}\mu_u(x)\,\mathrm{d}\varphi_\#\mu(u)\\
&= \int_{u\in B_+}\int_{y\in\psi^{-1}(\{t_B(u)\})} \zeta(u,t_B(u))\,\mathrm{d}\nu_{t_B(u)}(y)\,\mathrm{d}\varphi_\#\mu(u)\\
&= \int_{u\in B_+} \zeta(u,t_B(u))\,\mathrm{d}\varphi_\#\mu(u)\\
&= \int_{B_+\times B_-} \zeta(u,v)\,\mathrm{d}(\mathrm{id},t_B)_\#\varphi_\#\mu(u,v)\,,
\end{aligned}$$

proving the required equality and thus that T is a Monge map between μ and ν.

2.3 Proof of Theorem 3

We recall that we note $\mu' \triangleq \Phi_\#\mu$ and $\nu' \triangleq \Phi_\#\nu$. We also denote by B the image of $\varphi = p_B\circ\Phi$, so that μ',ν' are supported on $B\times F\subset B_0\times F$.

Step 1: Construction of the structured Monge map. Given that $\varphi_\#\mu$ is absolutely continuous w.r.t. the Lebesgue measure on the complete (separable) Riemannian manifold B_0, by Theorem 1 there exists a unique optimal transport plan $\pi_B^\star$ between $\varphi_\#\mu$ and $\varphi_\#\nu$ for the cost $\tilde{c}$ and it is induced by a map $t_B : B_0\to B_0$ of the form $t_B = \tilde{c}\text{-}\exp_u(\nabla f)$, with f $\tilde{c}$-convex.

By Lemma 2, we know that any transport plan $\pi\in\Pi(\mu,\nu)$ that satisfy $(\varphi,\varphi)_\#\pi = (\mathrm{id},t_B)_\#\mu$ must be optimal. Therefore, if π happens to be induced by a map T, that is $\pi = (\mathrm{id},T)_\#\mu$, we would obtain a Monge map between μ and ν. To build such a T, we proceed as in Section 2.2: we define a Monge map T_u between μ'_u and $\nu'_{t_B(u)}$ for $\varphi_\#\mu$-a.e. u and build a global map between μ' and ν' by (roughly) setting $T(u,x) = T_u(x)$. As in Section 2.2, proving the measurability of such T requires care.

Step 2: Transport between the fibers. For $\varphi_\#\mu$-a.e. u, μ'_u has a density w.r.t. the volume measure on F and the optimal cost between μ'_u and $\nu'_{t_B(u)}$ is finite by assumption. Whenever μ'_u has a density, we can therefore apply Proposition 2 between μ'_u and $\nu'_{t_B(u)}$ with the cost d_F^2 to obtain that there exists a plan π_u between these fibers that is induced by a map $T_u : F\to F$ that can be expressed as $T_u(v) = \exp_v(\nabla g_u(v))$ with g_u being $d_F^2/2$-convex on F.

Step 3: Measurability of the global map. Now that we have built structured maps T_u between corresponding fibers (through t_B), it remains to prove the existence of a measurable map $T : B_0\times F\to B_0\times F$ transporting μ' onto ν' satisfying $T(u,x) = (t_B(u),T_u(x))$ for $\varphi_\#\mu$-almost every u and μ'_u-almost every x. For this, we need an adaptation of Proposition 10 to the manifold setting. Namely, we have the following:

Proposition 11 (Measurable selection of maps, manifold case)**.** Let M be a complete Riemannian manifold and (B,Σ,m) a measure space. Consider a measurable function $B\ni u\mapsto(\mu_u,\nu_u)\in\mathcal{P}(M)^2$. Assume that for m-almost every $u\in B$, $\mu_u\ll\mathrm{vol}_M$ and μ_u and ν_u have a finite transport cost. Let T_u denote the (unique by Proposition 2) optimal transport map induced by the quadratic cost d_M^2 on M between μ_u and ν_u.

Then there exists a measurable function $(u, x) \mapsto T(u, x)$ such that m-a.e., $T(u, x) = T_u(x)$, μ_u-a.e.

This can essentially be shown by adapting the proof of [FGM10] to the manifold setting, and most steps adapt seamlessly. We provide a sketch of proof below. A complete proof, where we stress the points that need specific care in adaptation, is deferred to Appendix A.2.2.

Sketch of proof of Proposition 11. The map $(u, x) \mapsto T(u, x)$ is defined as being the limit of a Cauchy sequence (on a complete metric space) of measurable maps $(u, x) \mapsto T^{(k)}(u, x)$ which are built as L^1-approximations of the $T_u(x)$, yielding the relation $T(u, x) = T_u(x)$ for u, x-a.e. when $k \to \infty$.

More precisely, given a partition $(A_{n,k})_n$ of M made of cells of diameters $\leq 2^{-k}$, and given $a_{n,k} \in A_{n,k}$ any fixed point, $T^{(k)}$ is precisely defined as $T^{(k)}(u, x) = a_{n,k}$, where $a_{n,k}$ is the single point such that $T_u(x) \in A_{n,k}$. Since on each set of the form $\{(u, x) \mid T_u(x) \in A_{n,k}\}$, one has $d_M(T^{(k)}(u, x), T_u(x)) \leq 2^{-k}$ (the diameter of the cell), it follows that $T^{(k)}(u, x) \to T_u(x)$, at least for m-a.e. u and μ_u-a.e. x.

The key point is to ensure the measurability of the map $T^{(k)}$, which actually boils down to prove that the set $\{(u, x) \mid T_u(x) \in A_{n,k}\}$ is measurable. This is possible by relying on the theory of measurable sets-valued maps [RW09, Ch. 5 and 14], that we carefully adapt to the manifold setting (instead of $\mathbb{R}^d$) in Appendix A.2.1. □

We can apply this proposition with the manifold being the (common) fiber F on which the $\mu'_u, \nu'_{t_B(u)}$ are supported for $\varphi_\# \mu$-a.e. u, and for which we have access to the (unique) Monge map T_u. It gives the existence of a global map t_F satisfying $t_F(u, v) = T_u(v)$ for $\varphi_\# \mu$-a.e. u, and μ'_u-a.e. v, and we can thus define the (measurable) map $T(u, x) = (t_B(u), t_F(u, x))$.

An elementary computation using the pushforward and disintegration definitions as well as the facts that $T_{u\#}(\mu'_u) = \nu'_{t_B(u)}$ and $t_{B\#}(\varphi_\# \mu) = \varphi_\# \nu$ then guarantees that T sends μ' to ν', and $T_E \triangleq \Phi^{-1} \circ T \circ \Phi$ therefore sends μ to ν; and since $(\varphi, \varphi)_\#(\mathrm{id}, T_E)_\# \mu = \pi^\star_B$, by Lemma 2, T_E is an optimal map between μ and ν.

3 Applications to the quadratic and inner-product GW problems

In this section, we apply the results of Section 2 to our two problems of interest, (GW-IP) and (GW-Q), to obtain the existence of Monge maps as solutions of these GW problems in suitable settings. The proofs can be sketched in the following way: for either of these two problems,

- Let $\pi^\star$ be an optimal plan for the GW problem;
- Consider the associated linearized problem (1) at $\pi^\star$ which is a standard optimal transportation problem. From Section 1.1.2, we know that any solution of this OT problem ($\pi^\star$ being one of them) must be a solution of the initial GW problem;
- Observe that the linearized versions of (GW-IP) and (GW-Q) reduce to optimal transportation problems with costs that satisfy the assumptions of Theorem 3, thus guaranteeing the existence of a structured Monge map for the GW problem.

As such, we only focus on detailing this last point in the following subsections.

Remark 6. Our results provide the existence of a Monge map for the GW problems, but its construction depends on the linearization (that is, of the initial $\pi^\star$ selected). Thus, even though we would have uniqueness of solutions for the linearized problem as given *e.g.* by the subtwist condition, this does not give uniqueness of the solution to the corresponding GW problem: it may happen that the GW problem has several optimizers, each leading to a different linearized problem for which solutions, unique or not, may differ.

3.1 The inner-product cost

Expanding the integrand in (GW-IP) and using the fact that $\iint \langle x, x'\rangle^2 \,\mathrm{d}\pi \,\mathrm{d}\pi = \iint \langle x, x'\rangle^2 \,\mathrm{d}\mu \,\mathrm{d}\mu$ is constant (the same goes for the terms that depend on ν), one gets the equivalent problem

$$\min_{\pi\in\Pi(\mu,\nu)} \iint -\langle x, x'\rangle\langle y, y'\rangle \,\mathrm{d}\pi(x,y)\,\mathrm{d}\pi(x',y')\,.$$

This problem is not invariant to translations but it is to the action of $O_n(\mathbb{R}) \times O_d(\mathbb{R})$. Any optimal correspondence plan $\pi^\star$ is also an optimal transport plan for the linearized problem (1) with cost

$$C_{\pi^\star}(x,y) = -\int \langle x, x'\rangle\langle y, y'\rangle \,\mathrm{d}\pi^\star(x',y') = \left\langle -\int (y'\otimes x')x \,\mathrm{d}\pi^\star(x',y'),\, y\right\rangle = -\langle M^\star x,\, y\rangle\,,$$

where $M^\star \triangleq \int y'\otimes x' \,\mathrm{d}\pi^\star(x',y') \in \mathbb{R}^{d\times n}$. This linearized cost is twisted if and only if $M^\star$ is of full rank, hence in this case $\pi^\star$ is the only solution of the linearized problem and it is induced by a map, and Theorem 4.2.3 from [Vay20] gives a result on the structure of this map. We can actually generalize this to the case where $M^\star$ is arbitrary:

Theorem 4 (Existence of an optimal map for the inner product cost)**.** Let $n \geq d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)\times\mathcal{P}(\mathbb{R}^d)$ two measures with compact supports. Suppose that $\mu \ll \mathcal{L}^n$. Then there exists an optimal map for (GW-IP) that can be written as

$$T = O_{\mathcal{Y}}^\top \circ (T_0 \circ p_{\mathbb{R}^d}) \circ O_{\mathcal{X}}\,, \tag{4}$$

where $O_{\mathcal{X}}$ and $O_{\mathcal{Y}}$ are change-of-basis matrices of $\mathbb{R}^n$ and $\mathbb{R}^d$, $p_{\mathbb{R}^d} : \mathbb{R}^n \to \mathbb{R}^d$ is defined by $p_{\mathbb{R}^d}(x_1,\dots,x_n) = (x_1,\dots,x_d)$, and

$$T_0(x_1,\dots,x_d) = (\nabla f \circ \Sigma(x_1,\dots,x_h), \nabla g_{x_1,\dots,x_h}(x_{h+1},\dots,x_d))\,, \tag{5}$$

with $h \leq d$, $\Sigma \in \mathbb{R}^{h\times h}$ diagonal with positive entries, $f : \mathbb{R}^h \to \mathbb{R}$ convex and all $g_{x_1,\dots,x_h} : \mathbb{R}^{d-h} \to \mathbb{R}$ convex.

In order to show this, we will need two simple lemmas that essentially state that Brenier's result holds when the scalar product is perturbed by homeomorphisms. We state them now and prove them in Appendix A.1.

Lemma 3. Let $\mu, \nu \in \mathcal{P}(E)$ and let $\psi_1, \psi_2 : E \to F$ be homeomorphisms. Let $\tilde{c} : F\times F \to \mathbb{R}$ and consider the cost $c(x,y) = c(\psi_1(x), \psi_2(y))$. Then a map is optimal for the cost c between μ and ν if and only if it is of the form $\psi_2^{-1}\circ T\circ\psi_1$ with T optimal for the cost $\tilde{c}$ between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$.

The following lemma is a simple corollary of the first:

Lemma 4. Let $h \geq 1$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^h)$ with $\mu \ll \mathcal{L}^h$ with compact supports. Consider the cost $c(x,y) = -\langle\psi_1(x), \psi_2(y)\rangle$ where $\psi_1, \psi_2 : \mathbb{R}^h \to \mathbb{R}^h$ are diffeomorphisms. Then, there exists a unique optimal transport plan between μ and ν for the cost c, and it is induced by a map $t : \mathbb{R}^h \to \mathbb{R}^h$ of the form $t = \psi_2^{-1}\circ\nabla f\circ\psi_1$, with f convex.

We now prove Theorem 4.

Proof of Theorem 4. Using a singular value decomposition (SVD), we have $M^\star = O_{\mathcal{Y}}{}^\top \Sigma O_{\mathcal{X}} \in \mathbb{R}^{d\times n}$ with $O_{\mathcal{X}}, O_{\mathcal{Y}} \in O_n(\mathbb{R})\times O_d(\mathbb{R})$ orthogonal matrices of each Euclidean space and $\Sigma \in \mathbb{R}^{d\times n}$ diagonal with non-negative coefficients. The cost then becomes $C_{\pi^\star}(x,y) = -\langle O_{\mathcal{Y}}{}^\top\Sigma O_{\mathcal{X}} x,\, y\rangle = -\langle\Sigma(O_{\mathcal{X}}x),\, O_{\mathcal{Y}}y\rangle$. Using Lemma 3, the problem transforms into an optimal transportation problem between $\mu' \triangleq O_{\mathcal{X}\#}\mu$ and $\nu' \triangleq O_{\mathcal{Y}\#}\nu$; and choosing $O_{\mathcal{Y}}$ and $O_{\mathcal{X}}$ that sort the singular values in decreasing order, *i.e.* assuming $\sigma_1 \geq \cdots \geq \sigma_h > 0$ with $h \triangleq \mathrm{rk}(M^\star) \leq d$, the problem therefore transforms into $\min_{\tilde{\pi}}\langle c_\Sigma, \tilde{\pi}\rangle$ for $\tilde{\pi} \in \Pi(\mu',\nu')$, where $c_\Sigma(\tilde{x},\tilde{y}) = -\sum_{i=1}^h \sigma_i\tilde{x}_i\tilde{y}_i \triangleq \tilde{c}(p(\tilde{x}), p(\tilde{y}))$, p being the orthogonal projection on $\mathbb{R}^h$. We reduce to the case where both measures live in the same space by noting that since

$c_\Sigma(\tilde{x}, \tilde{y}) = c_\Sigma(p_{\mathbb{R}^d}(\tilde{x}), \tilde{y})$ for all $\tilde{x}$ and $\tilde{y}$, any map T_0 optimal between $\mu'' \triangleq p_{\mathbb{R}^d\#}\mu'$ and ν' will induce a map $T = T_0 \circ p_{\mathbb{R}^d}$ optimal between μ' and ν' (by Lemma 2). One can then recover an optimal map between μ and ν by composing with $O_{\mathcal{X}}$ and ${O_{\mathcal{Y}}}^\top$ (Lemma 3), hence Eq. (4).

The existence of such a map T_0 optimal between μ'' and ν' satisfying (5) follows from the application of Theorem 3 for $E = E_0 = \mathbb{R}^d = \mathbb{R}^h \times \mathbb{R}^{d-h} = B_0 \times F$ and $\varphi = p$. Indeed, B_0 and F are complete Riemannian manifolds; $\tilde{c}$ is twisted on $B_0 \times B_0$; $p_\#\mu''$ has a density on B_0 and every μ''_u has a density w.r.t. the Lebesgue measure on F as a conditional probability. We then make t_B explicit. One has that $c_\Sigma(x, y) = -\langle \tilde{\Sigma} x,\, y\rangle$, where $\tilde{\Sigma} = \operatorname{diag}(\sigma_i)_{1\le i\le h}$. As $p_\#\mu''$ has a density, we can apply Lemma 4 stated above with $(\psi_1, \psi_2) = (\tilde{\Sigma}, \mathrm{id})$ to obtain that there exists a unique optimal transport plan $\pi^\star_B$ between $p_\#\mu''$ and $p_\#\nu'$ for the cost c_Σ and that it is induced by a map $t_B : B \to B$ of the form $t_B = \nabla f \circ \tilde{\Sigma}$, with f convex. □

Remark 7. A particular case of our theorem is Theorem 4.2.3 from [Vay20] (Proposition 6 in this work): when $h = d$, the optimal map between $O_{\mathcal{X}\#}\mu$ and $O_{\mathcal{Y}\#}\nu$ writes $T_0 \circ p_{\mathbb{R}^d}$ with $T_0 = \nabla f \circ \tilde{\Sigma}$. The induced optimal map between μ and ν is then

$$T = O_{\mathcal{Y}}^\top \circ (\nabla f \circ \tilde{\Sigma} \circ p_{\mathbb{R}^d}) \circ O_{\mathcal{X}} = O_{\mathcal{Y}}^\top \circ (\nabla f \circ \Sigma) \circ O_{\mathcal{X}} = \nabla(f \circ O_{\mathcal{Y}}) \circ O_{\mathcal{Y}}^\top \circ \Sigma \circ O_{\mathcal{X}} = \nabla \tilde{f} \circ M^\star\,,$$

where $\tilde{f} \triangleq f \circ O_{\mathcal{Y}}$ is convex.

3.2 The quadratic cost

The (GW-Q) problem is invariant by translation of μ and ν. With no loss of generality, we suppose both measures centered. Expanding the integrand provides

$$\left||x - x'|^2 - |y - y'|^2\right|^2 = |x - x'|^4 + |y - y'|^4 - 2|x - x'|^2|y - y'|^2\,,$$

and the two first terms only depend on μ and ν, not on π. Expanding the remaining term yields nine terms. Two of them also lead to a constant contribution: $-|x|^2|y'|^2$ and $-|x'|^2|y|^2$; four lead to vanishing integrals since μ and ν are centered: $2|x|^2\langle y,\, y'\rangle$, $2|x'|^2\langle y,\, y'\rangle$, $2|y|^2\langle x,\, x'\rangle$ and $2|y'|^2\langle x,\, x'\rangle$. The remaining three terms then yield the following equivalent problem:

$$\min_{\pi\in\Pi(\mu,\nu)} \int -|x|^2|y|^2 \,\mathrm{d}\pi(x, y) + 2 \iint -\langle x,\, x'\rangle\langle y,\, y'\rangle \,\mathrm{d}\pi(x, y)\,\mathrm{d}\pi(x', y')\,.$$

Assuming an optimal correspondence plan $\pi^\star$, this plan is also an optimal transport plan for the linearized problem (1) with cost

$$C_{\pi^\star}(x, y) = -|x|^2|y|^2 - 4\int \langle x,\, x'\rangle\langle y,\, y'\rangle \,\mathrm{d}\pi^\star(x', y') = -|x|^2|y|^2 - 4\langle M^\star x,\, y\rangle\,,$$

where $M^\star \triangleq \int y' \otimes x' \,\mathrm{d}\pi^\star(x', y') \in \mathbb{R}^{d\times n}$. In the cases where the rank of $M^\star$ is d, this linearized cost satisfies both subtwist and 2-twist conditions, yielding an optimal bi-map that is also a map/anti-map, by compactness of the support of μ and ν and when μ has a density. Similarly, when the rank of $M^\star$ is $d - 1$, the cost only satisfies the 2-twist condition, yielding an optimal bi-map structure. In the case where $\operatorname{rk} M^\star \le d - 2$, nothing can be said and there is *a priori* no obvious reason for the existence of an optimal correspondence map; but perhaps surprisingly, it can actually be guaranteed.

Theorem 5 (Existence of an optimal map or bi-map for the quadratic cost)**.** Let $n \ge d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures with compact supports. Suppose that $\mu \ll \mathcal{L}^n$. Let $\pi^\star$ be a solution of (GW-Q) and $M^\star \triangleq \int y' \otimes x' \,\mathrm{d}\pi^\star(x', y')$. Then:

(i) if $\operatorname{rk} M^\star = d$, there exists an optimal plan that is induced by a bi-map which is also a map/anti-map;

(ii) if $\operatorname{rk} M^\star = d-1$, there exists an optimal plan that is induced by a bi-map;
(iii) if $\operatorname{rk} M^\star \leq d-2$, there exists an optimal plan that is induced by a map that can be written as

$$T = O_{\mathcal{Y}}{}^\top \circ T_0 \circ O_{\mathcal{X}}\,,$$

where $O_{\mathcal{X}}$ and $O_{\mathcal{Y}}$ are change-of-basis matrices of $\mathbb{R}^n$ and, writing any $x \in \mathbb{R}^n$ as $x = (x_H, x_\perp) \in \mathbb{R}^h \times \mathbb{R}^{n-h}$ and $\Phi(x) \triangleq (x_B, x_F) \triangleq ((x_H, |x_\perp|^2), x_\perp/|x_\perp|)$,

$$\Phi \circ T_0(x) = \Big(\tilde{c}\text{-}\exp_{x_B}(\nabla f(x_B)), \exp_{x_F}(\nabla g_{x_B}(x_F))\Big)$$

where explicit expressions for $\tilde{c}$ and $\tilde{c}$-exp are given in (6) and (7) respectively, with $h = \operatorname{rk} M^\star \leq d-2$, $f : \mathbb{R}^{h+1} \to \mathbb{R}$ being $\tilde{c}$-convex and all $g_{x_B} : \mathbb{R}^{n-h} \to \mathbb{R}$ being $d^2_{S^{n-h-1}}/2$-convex.

Proof.

(i) We show that in this case both subtwist and 2-twist conditions are satisfied. For subtwist, remark that $x \mapsto \nabla_x c(x, y_1) - \nabla_x c(x, y_2)$ has no solution if $|y_1| = |y_2|$ and a unique one otherwise. For 2-twist, remark that any $y \in \mathcal{Y}$ satisfying $(|y|^2 - |y_0|^2)x_0 = -2(M^\star)^\top(y - y_0)$ is fully determined by $r \triangleq |y|^2 - |y_0|^2$, and that r is solution of a polynomial of degree two. This proves that there are always at most two such y. Both subtwist and 2-twist conditions are therefore satisfied. The uniqueness of optimal plans given by subtwist then guarantees that the unique optimal transport plan for the linearized cost is induced both by a bi-map and a map/anti-map.

(ii) We show that in this case the 2-twist condition is satisfied. Let $y \in \mathcal{Y}$ such that $|y|^2 x_0 + (M^\star)^\top y = |y_0|^2 x_0 + (M^\star)^\top y_0 \triangleq v$. Up to singular value decomposition suppose $M^\star$ rectangular diagonal in $\mathbb{R}^{d\times n}$ with sorted singular values and write $\tilde{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_h)$ with $h \triangleq \operatorname{rk}(M^\star)$. Decompose each vector z of $\mathbb{R}^n$ or $\mathbb{R}^d$ as $z = (z_H, z_\perp)$, where $z_H \in \mathbb{R}^h$ and $z_\perp$ contains the remaining coordinates. The equation becomes:

$$\begin{cases} |y|^2 x_H + \tilde{\Sigma} y_H &= v_H \\ |y|^2 x_\perp &= v_\perp\,. \end{cases}$$

These two equations impose y to live in the intersection of a $(d-1)$-dimensional sphere and of a $(d-r)$-dimensional affine subspace of $\mathbb{R}^d$. As $r = d-1$, y belongs to a set of at most 2 points and the 2-twist condition is satisfied.

(iii) The case $\operatorname{rk} M^\star \leq d-2$ is a consequence of Theorem 3 and the proof is as follows. We consider the measure ν as a measure of $\mathbb{R}^n$ of d-dimensional support. Similarly to the inner product cost, by SVD the cost becomes $c(x, y) = -|x|^2|y|^2 - \langle O_{\mathcal{Y}}{}^\top \Sigma O_{\mathcal{X}} x,\, y\rangle = -|O_{\mathcal{X}} x|^2 |O_{\mathcal{Y}} y|^2 - \langle \Sigma(O_{\mathcal{X}} x),\, O_{\mathcal{Y}} y\rangle$, and using Lemma 3 the problem transforms into $\min_{\tilde{\pi}} \langle c_\Sigma, \tilde{\pi}\rangle$ for $\tilde{\pi} \in \Pi(O_{\mathcal{X}\#}\mu, O_{\mathcal{Y}\#}\nu)$, where $c_\Sigma(x, y) \triangleq -|x|^2|y|^2 - \langle \Sigma x,\, y\rangle$. Further assuming $\sigma_1 \geq \cdots \geq \sigma_h > 0$ and writing any $z \in \mathbb{R}^n$ as $z = (z_H, z_\perp) \in \mathbb{R}^h \times \mathbb{R}^{n-h}$,

$$\begin{aligned} c_\Sigma(x, y) &= -|x_H|^2|y_H|^2 - |x_H|^2|y_\perp|^2 - |x_\perp|^2|y_H|^2 - |x_\perp|^2|y_\perp|^2 - \langle \tilde{\Sigma} x_H,\, y_H\rangle \\ &= -|x_H|^2|y_H|^2 - |x_H|^2 y_+ - x_+|y_H|^2 - x_+ y_+ - \langle \tilde{\Sigma} x_H,\, y_H\rangle \\ &\triangleq \tilde{c}(\varphi(x), \varphi(y))\,, \end{aligned} \tag{6}$$

with $x_+ = |x_\perp|^2$, $y_+ = |y_\perp|^2$ and $\varphi : x \mapsto (x_H, |x_\perp|^2)$, and the cost $c_\Sigma(x, y)$ only depends of the values of $\varphi(x)$ and $\varphi(y)$. A direct computation gives that $\tilde{c}$ satisfies the twist condition.
Now, the same as in Example 2 applies, but this time with $E_0 = \mathbb{R}^h \times \mathbb{R}^{n-h}$, $E = E_0 \backslash (\mathbb{R}^h \times \{0\})$, $B_0 = \mathbb{R}^h \times \mathbb{R}$ and $F = S^{n-h-1} = \{x \in E_0 \mid |x_\perp| = 1\}$. Is then ensured the existence of a structured Monge map between μ and ν for the cost c: it decomposes for almost all $x \in \mathbb{R}^n$ as a Monge map on the basis $B_0 = \mathbb{R}^{h+1}$ obtained as the gradient of a $\tilde{c}$-convex function $f : \mathbb{R}^{h+1} \to \mathbb{R}$ (via the $\tilde{c}$-exponential map on $\mathbb{R}^{h+1}$) and a Monge map on each fiber $F = S^{n-h-1}$, also built from

gradients of convex functions $h_{(x_H,|x_\perp|^2)} : S^{n-h-1} \to \mathbb{R}$ (via the exponential map on the sphere); hence the result. Note that the $\tilde{c}$-exponential map is given in closed form by

$$\tilde{c}\text{-}\exp_u(p) = \begin{pmatrix} \tilde{\Sigma}^{-1}(p_H - 2p_+u) \\ p_+ - |\Sigma^{-1}(p_H - 2p_+u)|^2 \end{pmatrix} \quad \text{for all } u \in \mathbb{R}^h \text{ and } p = (p_H, p_+) \in \mathbb{R}^h \times \mathbb{R}\,. \tag{7}$$

Last, note that the case where $M^\star = 0$ has not been explicitly treated. In this case, the cost is simply $c(x,y) = -|x|^2|y|^2 = \tilde{c}(x_+, y_+)$ and the strategy above directly applies. □

3.3 Complementary study of the quadratic cost in the one-dimensional case

The (GW-Q) problem being invariant by translation, we assume that measures μ and ν below are centered. In the one-dimensional case $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$, the linearized GW problem (1) reads, with $\pi^\star$ an optimal correspondence plan:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} (-x^2y^2 - 4mxy)\,\mathrm{d}\pi(x,y)\,, \quad \text{where } m = \int_{\mathcal{X}\times\mathcal{Y}} x'y'\,\mathrm{d}\pi^\star(x',y')\,, \tag{8}$$

and for any plan $\pi \in \Pi(\mu,\nu)$ (not necessarily optimal), we denote by $m(\pi) = \int xy\,\mathrm{d}\pi(x,y)$ what we call the *correlation* of π. When $m \neq 0$ (resp. $m = 0$), the associated OT cost function $c_m(x,y) = -x^2y^2 - 4mxy$ satisfies both subtwist and 2-twist conditions (resp. satisfies the 2-twist condition) and therefore induces an optimal transport plan that is a bi-map and a map/anti-map (resp. a bi-map), see Theorem 5. In the following sections, we study the tightness of this result, asking if it exists cases where the optimal plan for this cost is not a map. In the one-dimensional case, the *submodularity* property of the cost c is useful, sometimes called the Spence–Mirrlees condition. It guarantees the optimality of the non-decreasing (resp. non-increasing) matching $\pi^\oplus_{\text{mon}}$ (resp. $\pi^\ominus_{\text{mon}}$) [Car12, San15]:

Definition 4 (Submodular cost)**.** *Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$ be two intervals. A function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is said to be* submodular *if for all $(x,y) \in \mathcal{X}, \mathcal{Y}$ and $\delta_1, \delta_2 \in \mathbb{R}_{\geq 0}$ such that $x + \delta_1 \in \mathcal{X}$ and $y + \delta_2 \in \mathcal{Y}$,*

$$c(x,y) + c(x+\delta_1, y+\delta_2) \leq c(x, y+\delta_2) + c(x+\delta_1, y)\,.$$

If c is twice differentiable, this global condition on the rectangle $\mathcal{X} \times \mathcal{Y}$ is equivalent to the local condition

$$\text{for all } (x,y) \in \mathcal{X} \times \mathcal{Y}, \quad \partial_{xy} c(x,y) \leq 0\,. \tag{Submod}$$

Supermodularity *is defined with the reversed inequality.*

We state the well-known consequence of submodularity on the structure of optimal plans.

Proposition 12. Let $\mu, \nu \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ of finite transport cost. If c satisfies (Submod), then $\pi^\oplus_{\text{mon}}$ is an optimal plan for (OT), unique if the inequality is strict everywhere on $\mathcal{X} \times \mathcal{Y}$ (see [McC12]). Similarly, supermodularity induces the optimality of $\pi^\ominus_{\text{mon}}$.

The linearized quadratic GW cost with parameter $m \geq 0$ is submodular on the region $S = \{(x,y) \mid xy \geq -m\}$ and supermodular elsewhere (see Figure 2 for an illustration); so we cannot directly apply this proposition. In general, for a given cost in dimension one, there are regions of sub and supermodularity and this property gives only few information without particular structures of the regions. To be applicable, submodularity requires a rectangular region. In view of the form of the regions of modularity in our particular case, we can state:

Proposition 13. Let $m > 0$, $S = \{(x,y) \mid xy \geq -m\}$ and denote by $\pi^\star$ an optimal transportation plan for the cost c_m. Then, $[\pi^\star]_{|S}$ (the plan restricted to the submodularity region) is the monotone non-decreasing rearrangement between its marginals. Similarly, if $m < 0$, $[\pi^\star]_{|\bar{S}}$ (the plan restricted to the supermodularity region) is the monotone non-increasing rearrangement. If $m = 0$, there are four regions of sub/supermodularity, and on each of these regions, the plan is a monotone rearrangement.

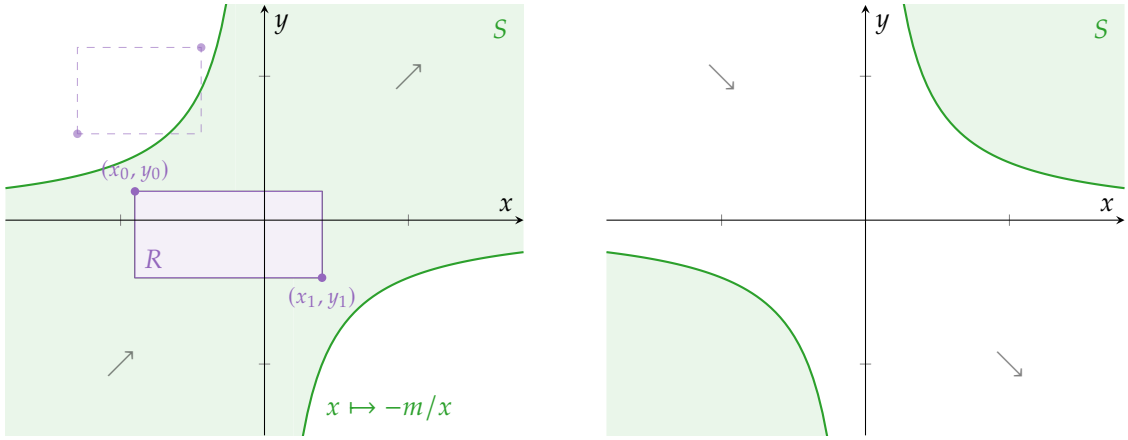


Figure 2: Submodularity region S (light green) and supermodularity region $\overline{S}$ for the linearized quadratic GW cost with parameter $m > 0$ **(Left)** or $m < 0$ **(Right)**, with an example of a rectangle $R \subset S$ (purple) defined by a decreasing matching (x_0, y_0), (x_1, y_1) that allows to conclude for the monotonicity of the optimal plan on S (see proof of Proposition 13), while the same argument does not apply to rectangles in $\overline{S}$ (light purple), see Remark 8. Increasing ($\nearrow$) and decreasing ($\searrow$) arrows are drawn in regions where the monotonicity of optimal correspondence plans is given by Proposition 13.

Figure 2 illustrates the regions, with increasing ($\nearrow$) and ($\searrow$) decreasing arrows, where the monotonicity of optimal correspondence plans is given by the proposition.

Proof. In this proof, we use the fact that if $\pi^\star$ is optimal, then it is also optimal when restricted on a domain between the corresponding marginals. In particular, when $m > 0$, the plan $[\pi^\star]_{|S}$ is necessarily optimal for the cost c_m between its marginals. Consider $(x_0, y_0), (x_1, y_1) \in \mathrm{supp}([\pi^\star]_{|S})$ such that $x_0 < x_1$ and $y_0 > y_1$. The particular geometry of the space S ensures that in this particular configuration, the rectangle $R = [x_0, x_1] \times [y_1, y_0]$ is necessarily contained in S (see Figure 2). As a consequence, the strict submodularity of the cost applies on R to prove that $[\pi^\star]_{|R}$ is the monotone non-decreasing plan, contradicting the configuration. The proof goes similarly when $m < 0$. The case $m = 0$ yields four different regions of sub/supermodularity which are quadrants. Since these regions are union of rectangles on which the cost is strictly sub/supermodular, the argument above applies. □

Remark 8. Let us underline that it is not possible to apply the argument to the whole region S since any two points of S are not contained in a rectangle included in S. The proof above only uses the fact that this property is used for some particular configuration of points, namely whenever $x_0 < x_1$ and $y_0 > y_1$, which is (somewhat fortunately) exactly the configuration we want to consider when wondering if a map may be non-increasing. For instance, this argument does not apply to the complementary of S denoted $\overline{S}$, although the property is of course satisfied on every rectangle contained in $\overline{S}$.

Before going into further details on our complementary study, we recall the discrete formulation of (OT) in dimension one. Given two sets $\{x_1, \dots, x_N\}$ and $\{y_1, \dots, y_M\}$ of $\mathbb{R}$ and two probability vectors a and b, the (OT) problem between the discrete measures $\mu = \sum_{i=1}^N a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^M b_j \delta_{y_j}$ reads

$$\min_{\pi \in U(a,b)} \langle C, \pi \rangle ,$$

where $U(a, b) \triangleq \{\pi \in \mathbb{R}^{N \times M} \mid \pi \mathbf{1}_M = a,\ \pi^\top \mathbf{1}_N = b\}$ is the *transport polytope*, $C = (c(x_i, y_j))_{i,j}$ is the cost matrix and $\langle \cdot, \cdot \rangle$ is the Frobenius inner product. In the case of the linearized problem (8), we denote

by $C_{\mathrm{GW}(m)}$ the cost matrix, that has coefficients $(C_{\mathrm{GW}(m)})_{i,j} = -x_i^2 y_j^2 - 4m x_i y_j$ with $m = \langle C_{xy}, \pi^\star \rangle$ and $(C_{xy})_{i,j} = x_i y_j$.

In the following sections, we study the optimality of the monotone non-decreasing and non-increasing rearrangements $\pi^{\oplus}_{\mathrm{mon}}$ and $\pi^{\ominus}_{\mathrm{mon}}$. It is worth noting that by submodularity of $x, y \mapsto -xy$, these two plans have respective correlations $m_{\min}$ and $m_{\max}$, where

$$\begin{cases} m_{\min} & = \min_\pi \langle C_{xy}, \pi \rangle \\ m_{\max} & = \max_\pi \langle C_{xy}, \pi \rangle \end{cases}, \quad \text{with } (C_{xy})_{i,j} = x_i y_j \,, \tag{9}$$

and that for any correspondence plan π, the value of its correlation $m(\pi)$ lies in the interval $[m_{\min}, m_{\max}]$. We provide in the following a complementary study of the quadratic cost in dimension one, namely

(i) a procedure to find counter-examples to the optimality of the monotone rearrangements;
(ii) empirical evidence for the tightness of Theorem 5;
(iii) a proof of the instability of having a monotone rearrangement as an optimal correspondence plan;
(iv) a new result on the optimality of the monotone rearrangements when the measures are composed of two distant parts.

All experiments are reproducible and the code can be found on GitHub[2].

3.3.1 Adversarial computation of non-monotone optimal correspondence plans. Theorem 4.1.1 of [Vay20] claims that in the discrete case in dimension 1 with $N = M$ and $a = b = \mathbf{1}_N$, the optimal solution of (QAP) is either the monotone non-decreasing rearrangement $\pi^{\oplus}_{\mathrm{mon}}$ or the monotone non-increasing one $\pi^{\ominus}_{\mathrm{mon}}$. It seems to be the case with high probability empirically when generating random discrete measures. While this claim is true for $N = 1, 2$ and 3, a counter-example for $N \geq 7$ points has recently been exhibited in [BHS22]. We further propose a procedure to automatically obtain additional counter-examples, demonstrating empirically that such adversarial distributions occupy a non-negligible place in the space of empirical measures. We propose to perform a gradient descent over the space of empirical distributions on $\mathcal{X} \times \mathcal{Y}$ using an objective function that favors the strict sub-optimality of the monotone rearrangements; we now detail this procedure.

For $N \geq 1$, we consider the set of empirical distributions over $\mathcal{X} \times \mathcal{Y} = \mathbb{R} \times \mathbb{R}$ with N points and uniform mass, *i.e.* of the form $\pi = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}$. Such plans π can be seen as the identity mapping between vectors $X = (x_1, \dots, x_N)$ and $Y = (y_1, \dots, y_N)$, and we therefore note $\pi = \mathrm{id}(X, Y)$. Denoting by c_{GW} the functional that takes a correspondence plan and returns its cost on the GW problem, we then define $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ by

$$\mathcal{F}(X, Y) \triangleq c_{\mathrm{GW}}(\pi) - \min \left\{ c_{\mathrm{GW}}(\pi^{\oplus}_{\mathrm{mon}}), \, c_{\mathrm{GW}}(\pi^{\ominus}_{\mathrm{mon}}) \right\},$$

where $\pi = \mathrm{id}(X, Y)$ and $\pi^{\oplus}_{\mathrm{mon}}$ and $\pi^{\ominus}_{\mathrm{mon}}$ are the monotone rearrangements between X and Y. This quantifies how well the plan π compares to the two monotone rearrangements. We generate N points at random in $[0, 1]^2$ and then perform a simple gradient descent over the positions of the points $(X, Y) = (x_i, y_i)_i$ following the objective

$$\min_{X, Y \in \mathbb{R}^N} \mathcal{F}(X, Y) \,.$$

We include an early-stopping threshold t since when $\mathcal{F}(\pi)$ becomes negative, *i.e.* we found an adversarial example, the objective function often decreases exponentially fast. The procedure can be found in Algorithm 1 below. We implemented it using PyTorch's automatic differentiation [PGM+19] and used [BTBD20] to implement a differentiable sorting operator to compute the monotone rearrangements. Adversarial plans $\pi_f = \mathrm{id}(X_f, Y_f)$ obtained by Algorithm 1 are not *a priori* optimal for the GW cost between their marginals; but they have at least a better cost than the monotone rearrangements since $\mathcal{F}(X_f, Y_f) < 0$, proving the sub-optimality of the latter.

[2]link of the code: https://github.com/theodumont/monge-gromov-wasserstein.

Algorithm 1 Simple gradient descent over the positions $(x_i)_i$ and $(y_i)_i$.

Parameters:

- N: number of points of the distributions
- N_{iter}: maximum number of iterations
- η: step size
- t: early stopping threshold

Algorithm:

$X \leftarrow N$ random values in $[0,1]$, then centered
$Y \leftarrow N$ random values in $[0,1]$, then centered
for $i \in \{1, \dots, N_{\text{iter}}\}$ **do**
  $\pi^{\oplus}_{\text{mon}} \leftarrow \texttt{id(sort}(X)\texttt{,sort}(Y)\texttt{)}$ ▹ `id` is the identity mapping
  $\pi^{\ominus}_{\text{mon}} \leftarrow \texttt{id(sort}(X)\texttt{,reverse(sort}(Y)\texttt{))}$
  $\pi \leftarrow \texttt{id}(X,Y)$
  $\mathcal{F}(X,Y) \leftarrow \texttt{GW}(\pi) - \texttt{min(GW}(\pi^{\oplus}_{\text{mon}})\texttt{,GW}(\pi^{\ominus}_{\text{mon}})\texttt{)}$ ▹ `GW` computes c_{GW}
  if $\mathcal{F}(X,Y) < t$ **then** stop ▹ early stopping
  $(X,Y) \leftarrow (X,Y) - \eta \nabla \mathcal{F}(X,Y)$ ▹ step of gradient descent
end for
return $\pi_f = \texttt{id}(X,Y)$

Output: a plan π_f with better GW cost than $\pi^{\oplus}_{\text{mon}}$ and $\pi^{\ominus}_{\text{mon}}$

On Figure 3 is displayed an example of adversarial plans obtained following this procedure. It can be observed that during the descent, the plan π has difficulties getting out of what seems to be a saddle point consisting in being (close to) the monotone rearrangements between its marginals. Moreover, it is worth noting that the marginals of our typical adversarial plans, such as the one of Figure 3, are often similar to the counter-example proposed in [BHS22], where both measures have their mass concentrated near zero, except for one outlier for μ and two for ν, one on each tail.

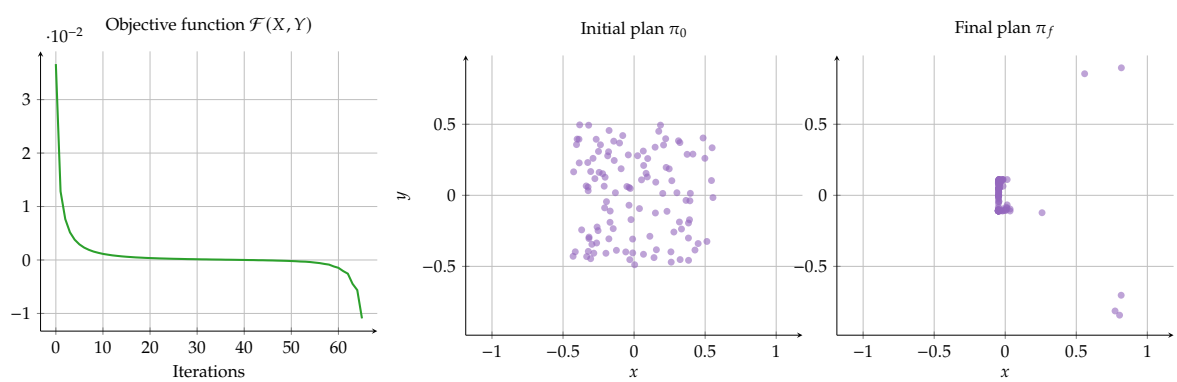


Figure 3: Gradient descent results with parameters $N = 122$, $\eta = 26$, $t = -2$. **(Left)** Evolution of the objective function $\mathcal{F}$. **(Center)** Initial plan π_0, generated at random. **(Right)** Final plan π_f (iter. 66).

Furthermore, examining the optimal correspondence plan for these adversarial examples allows to exhibit cases where it is not a map, providing empirical evidence for the following conjecture:

Conjecture 1. Theorem 5 is tight, *i.e.* there exists μ and ν for which optimal correspondence plans for (GW-Q) are not maps but rather a union of two graphs (that of two maps or one map and one

anti-map); and this even if μ has a density, classical OT assumption for the existence of an optimal transport map.

In order to approximate numerically the case of a measure which has density w.r.t. the Lebesgue measure, we convolve our distributions $\mu = (X_f, \mathbf{1}_N)$ and $\nu = (Y_f, \mathbf{1}_N)$ with a Gaussian of standard deviation σ and represent it in Eulerian coordinates; that is we evaluate the closed form density on a fine enough grid. When σ is large, the optimal correspondence plan for GW is probably induced by a monotone map, as it is the case very frequently empirically; on the contrary, if σ is sufficiently small, *i.e.* when the distributions are very close to their sum of Dirac measures discrete analogous, the optimal correspondence plan should not be a monotone map, by construction of μ and ν.

Remark 9. Because of the adversarial nature of π_f for the sub-optimality of $\pi^{\oplus}_{\text{mon}}$ and $\pi^{\ominus}_{\text{mon}}$, we know that when σ is sufficiently small, the optimal correspondence plan is not a monotone rearrangement. Still, it could be the case that this optimal plan is a map, but not a monotone one, and there is *a priori* no reason to believe that π_f will agree with Conjecture 1. Surprisingly, it sometimes does, as numerical experiments below suggest.

In order to find the optimal correspondence plan $\pi^\star$ between μ and ν, we leverage the fact that $\pi^\star$ is a solution of its associated linearized problem. Therefore, a minimizer of the GW functional is given by

$$\arg\min \left\{ \mathrm{GW}(\pi^\star_m) \,\middle|\, \pi^\star_m \in \operatorname*{arg\,min}_{\pi \in \mathcal{U}(a,b)} \langle C_{\mathrm{GW}(m)}, \pi \rangle \,,\, m \in [m_{\min}, m_{\max}] \right\}, \tag{10}$$

where $(C_{\mathrm{GW}(m)})_{i,j} = -x_i^2 y_j^2 - 4m x_i y_j$. We therefore compute both $m_{\min}$ and $m_{\max}$ by solving the linear programs in (9), discretize the interval $[m_{\min}, m_{\max}]$ with $N_{\Delta m}$ points, and solve the corresponding linear optimization problem for every value of the parameter m and evaluate the GW cost on each optimal plan for the given parameter m. We then check if the optimal plan exhibits a bi-map structure. The procedure is described in Algorithm 2.

We display the results on Figure 4, where we plot the optimal correspondence plan $\pi^\star$ in two cases:

(a) starting from an adversarial plan with both marginals convolved as to simulate densities;
(b) starting from an adversarial plan with only the first marginal convolved and the second marginal being a sum of Dirac measures.

To facilitate the reading, we draw a blue pixel at a location x on the discretized x-axis (resp. y on the y-axis) each time x (resp. y) has two (disjoint) images (resp. pre-images), making $\pi^\star$ a bi-map (resp. a bi-anti-map), or the union of a graph and an anti-graph. In both cases, we observe that $\pi^\star$ is not a map but a bi-map instead, similarly to [CMN10, Sec. 4.5]. Note that in case (b), ν being atomic, there cannot be a map from ν to μ, so in both (a) and (b) we numerically exhibit an instance where there is *a priori* no map from neither μ to ν nor ν to μ. We also plot the submodularity regions of the linearized GW cost function with parameter $m(\pi^\star)$ as an overlay and we observe that when the optimal plan gives mass to a region where the cost is submodular (resp. supermodular), it has a monotone non-decreasing (resp. non-increasing) behaviour in this region.

3.3.2 Empirical instability of the optimality of monotone rearrangements. The above study demonstrates that there exist probability measures μ and ν for which property

$$P(\mu, \nu): \quad \pi^{\oplus}_{\text{mon}} \text{ or } \pi^{\ominus}_{\text{mon}} \text{ is an optimal correspondence plan between } \mu \text{ and } \nu$$

does not hold. However, as it is very likely in practice when generating empirical distributions at random, one could ask if property P is at least *stable*, *i.e.* if when we have μ_0 and ν_0 satisfying $P(\mu_0, \nu_0)$ there is a small ball around μ_0 and ν_0 (for a given distance, say Wasserstein-2) inside which property P remains valid. A negative answer to this—besides, in the symmetric case—is given by the counter-example by [BHS22] with an increasing number of points:

Proposition 14. There exists two *symmetric* measures μ, ν on $\mathbb{R}$ and sequences $(\mu_n)_n$, $(\nu_n)_n$ that weakly converge to μ, ν such that optimal plans π_n between μ_n and ν_n are never supported by a monotone

Algorithm 2 Generating bi-maps from adversarial examples.

Input: an adversarial plan $\pi_f = \mathrm{id}(X_f, Y_f)$ obtained from Algorithm 1

Parameters:
- σ: standard deviation of convolution
- $N_{\Delta x}$: discretization precision
- $N_{\Delta m}$: discretization precision of the interval $[m_{\min}, m_{\max}]$

Algorithm:

$a \leftarrow$ `convolution`$(X_f, \sigma, N_{\Delta x})$
$b \leftarrow$ `convolution`$(Y_f, \sigma, N_{\Delta x})$ ▹ (optional)
$m_{\min} \leftarrow \min_{\pi \in \mathcal{U}(a,b)} \langle C_{xy}, \pi \rangle$ ▹ solve linear programs
$m_{\max} \leftarrow \max_{\pi \in \mathcal{U}(a,b)} \langle C_{xy}, \pi \rangle$
`scores` $\leftarrow$ `[]`
for $m \in \{m_{\min}, \dots, m_{\max}\}$ **do** ▹ with $N_{\Delta m}$ points
  $\pi^\star_m \leftarrow \arg\min_{\pi \in \mathcal{U}(a,b)} \langle C_{\mathrm{GW}(m)}, \pi \rangle$ ▹ solve linear program
  append $\mathrm{GW}(\pi^\star_m)$ to `scores`
end for
$\pi^\star \leftarrow \arg\max_\pi$ `scores` ▹ take best plan for GW
$b \leftarrow$ "$\pi^\star$ is a bi-map"
return $\pi^\star, b$

Outputs:
- $\pi^\star$: optimal plan for GW
- b: boolean asserting if $\pi^\star$ is a bi-map

map.

Proof. We consider $\mu = \nu = \delta_0$ and the discrete measures $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ defined as follows for $n \geq 7$:

$$x_i \triangleq \begin{cases} -1 & \text{for } i = 1 \\ (i - \frac{n+1}{2})\frac{1}{n^2} & \text{for } i = 2, \dots, n-1 \\ 1 & \text{for } i = n \end{cases} \quad \text{and} \quad y_i \triangleq \begin{cases} -1 & \text{for } i = 1 \\ -1 + \frac{1}{n^2} & \text{for } i = 2 \\ (i-2)\frac{1}{n^2} & \text{for } i = 3, \dots, n \end{cases}$$

which is simply the counter-example from [BHS22] with n points and $\varepsilon_n = 1/n^2$. Since $n \geq 7$, $\varepsilon_n < 2/(n-3)$ and the identity or anti-identity mappings are not optimal between μ_n and ν_n. By direct computation,

$$\mathrm{W}_2^2(\delta_0, \mu_n) = O(2/n + \varepsilon_n^2 n^2) \xrightarrow{n \to \infty} 0\,,$$

and the exact same goes for $\nu = \delta_0$ and ν_n. □

One can actually obtain non-degenerate (although not symmetric anymore) examples of such measures μ, ν. We start from the counter-example given in [BHS22] with $N = 7$ points and $\varepsilon = 10^{-2}$, that we convolve with a Gaussian of standard deviation σ as before. We then plot as a function of $m \in [m_{\min}, m_{\max}]$ the (true) GW cost of a plan $\pi^\star_m$ optimal for the linearized GW problem: $\pi^\star_m \in \arg\min_\pi \langle C_{\mathrm{GW}(m)}, \pi \rangle$. The minimum values of this graph are attained by the correlations of optimal correspondence plans, as explained in Section 3.3.1. Hence if σ is small, this optimal plan is not a monotone rearrangement by construction and the minima are not located on the boundary of the domain. On the contrary, when σ is large, a monotone rearrangement is optimal again. In order

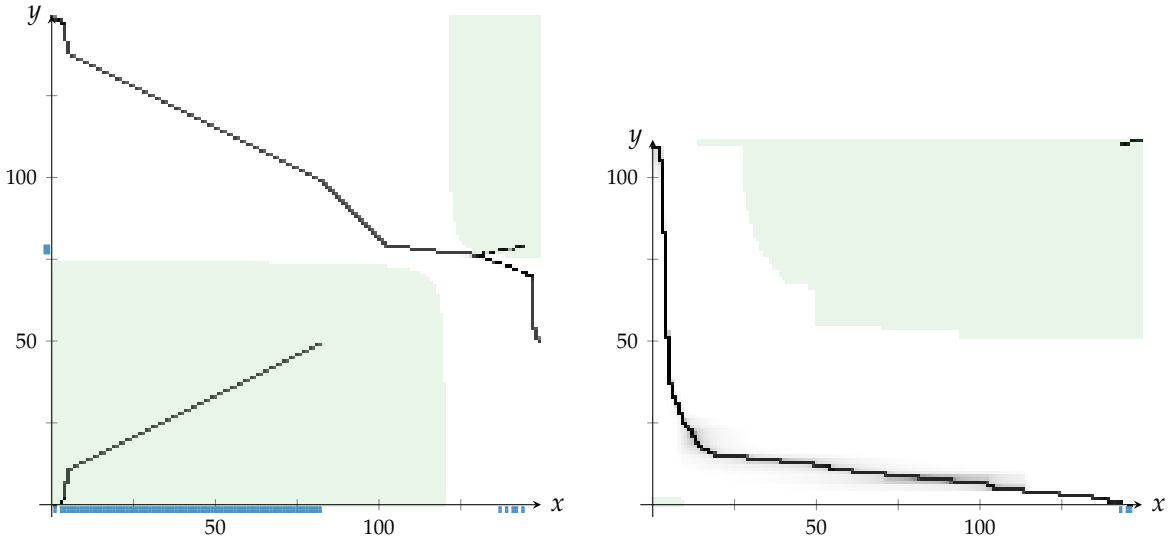
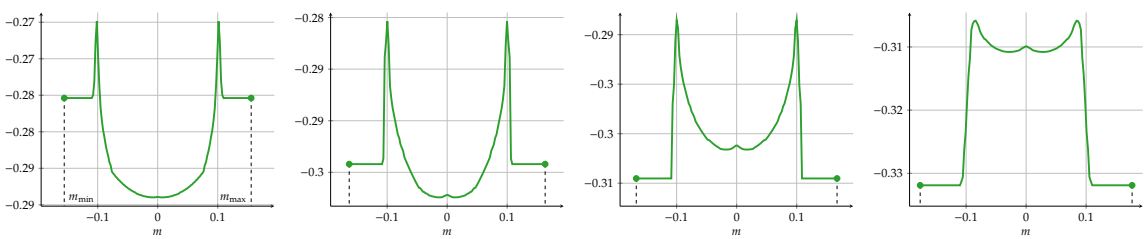


Figure 4: Optimal correspondence plan (in log scale) obtained with our procedure, starting either from a plan with both marginals convolved **(Left)** or with only the first marginal convolved **(Right)**; bi-map and anti-bi-map coordinates (blue); submodularity regions (light green). Parameters: $\sigma = 5.10^{-3}$, $N_{\Delta x} = 150$, $N_{\Delta m} = 2000$.



$\sigma_1 = 8.10^{-3}$ $\sigma_2 = 8.8.10^{-3}$ $\sigma_3 = 10^{-2}$ $\sigma_4 = 3.10^{-2}$

Figure 5: Evolution of the graph of $m \mapsto \mathrm{GW}(\pi^\star_m)$ when varying σ on the counter-example of [BHS22] with $N = 7$ points and $\varepsilon = 10^{-2}$. Parameters: $N_{\Delta x} = 100$, $N_{\Delta m} = 150$.

to study the phase transition, we plot on Figure 5 the landscape of $m \mapsto \mathrm{GW}(\pi^\star_m)$ while gradually increasing the value of σ.

Looking at Figure 5, it is worth noting that there is an incentive for optimal plans at correlations close to $m_{\min}$ or $m_{\max}$ to be the monotone rearrangements, as the horizontal portions of the plot suggest. More importantly, it can be observed that when $\sigma = \sigma_3$ or σ_4, the monotone rearrangements are optimal, as their correlations realize the minimum of $m \mapsto \mathrm{GW}(\pi^\star_m)$; unlike for σ_1 and σ_2, for which the minimum value of the plot is located near zero. Hence there should exist a $\sigma_0 \in (\sigma_2, \sigma_3)$ for which the convolved measures have both $\pi^{\oplus}_{\mathrm{mon}}$, $\pi^{\ominus}_{\mathrm{mon}}$ and another π_0 as optimal correspondence plans; it is direct that property P does not hold in the neighbourhood of these specific measures μ_0 and ν_0.

3.3.3 A positive result for measures with two components. In the following, μ_1, μ_2, ν_1 and ν_2 are four probability measures supported on a compact interval $A \subset \mathbb{R}$. Denote $\Delta = \mathrm{diam}(A)$, and fix $t \in (0,1)$ and $K > \Delta$. Let $\tau_K : x \mapsto x+K$ denote the translation by K, and $A+K = \tau_K(A) = \{x+K \mid x \in A\}$.

Now, introduce the measures

$$\mu = (1-t)\mu_1 + t\tau_{K\#}\mu_2 \quad \text{and} \quad \nu = (1-t)\nu_1 + t\tau_{K\#}\nu_2 \,.$$

Note that μ_1 and $\tau_{K\#}\mu_2$ (resp. ν_1 and $\tau_{K\#}\nu_2$) have disjoint supports. We want to prove the following:

Proposition 15. For K large enough, the unique optimal plan for the quadratic cost between μ and ν is given by one of the two monotone maps (non-decreasing or non-increasing).

Remark 10. The hypothesis of the theorem illustrates that monotone maps are favored when μ and ν both contain a single or more outliers. The proof of the theorem actually shows the importance of long range correspondences or global effect over the local correspondences on the plan. In other words, even though monotone maps may not be optimal locally, global correspondences favor them. Moreover, these global correspondences have proportionally more weight in the GW functional since the cost is the squared difference of the squared distances. In conclusion, pair of points which are at long distances tend to be put in correspondence. In turn, this correspondence, as shown in the proof, favors monotone matchings. This argument gives some insight on the fact that a monotone map is often optimal and it is made quantitative in the previous theorem.

We first prove the following lemma:

Lemma 5. In the setting described above, there exists $K_0 > 0$ such that if $K \geq K_0$, every π optimal plan for $\mathrm{GW}(\mu, \nu)$ can be decomposed as $\pi = \pi_1 + \pi_2$, where either:

1. π_1 is supported on $A \times A$ and π_2 on $(A+K) \times (A+K)$ (that is, we separately transport μ_1 to ν_1 and $\tau_{K\#}\mu_2$ to $\tau_{K\#}\nu_2$), or
2. π_1 is supported on $A \times (A+K)$ and π_2 on $A \times (A+K)$ (that is, we transport μ_1 to $\tau_{K\#}\nu_2$ and μ_2 to $\tau_{K\#}\nu_1$).

Furthermore, whenever $t \neq \frac{1}{2}$, only the first point can occur.

Proof. Consider first the case $t = \frac{1}{2}$. To shorten the notations, we introduce the notations $A_1 = A$ and $A_2 = A + K$. We can now decompose any plan π as $\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22}$ where for instance π_{12} denotes the restriction of the plan π to the product $A_1 \times A_2$. Let us also denote by r the mass of π_{12}, one has $0 \leq r \leq 1/2$ and by symmetry, one can choose that $r \leq 1/4$, otherwise we exchange A_1 and A_2 for the second measure since the cost is invariant to isometries. Remark that, due to marginal constraints, the total mass of π_{11} and π_{22} is $1/2 - r$ and the mass of π_{21} is r. Therefore, it is possible to consider a coupling plan $\tilde{\pi}_{11}$ between the first marginal of π_{12} and the second marginal of π_{21}, and similarly, let $\tilde{\pi}_{22}$ be a coupling plan between the first marginal of π_{21} and the second marginal of π_{12}. We then define a competitor plan $\tilde{\pi} = \pi_{11} + \tilde{\pi}_{11} + \pi_{22} + \tilde{\pi}_{22}$. The first step is to get a lower bound on the term $\mathrm{GW}(\pi, \pi)$. Slightly overloading the notations, we introduce

$$\mathrm{GW}(\pi, \gamma) = \int c \, \mathrm{d}\pi \otimes \gamma \,. \tag{11}$$

We expand GW by bilinearity

$$\mathrm{GW}(\pi, \pi) = \sum_{i,j,i',j'} \mathrm{GW}(\pi_{ij}, \pi_{i'j'}) = \sum_{i,j} \mathrm{GW}(\pi_{ii}, \pi_{jj}) + R \,,$$

where R is the remainder that contains 12 terms from which one can identify two types. Among them, 8 terms are of the type $\mathrm{GW}(\pi_{12}, \pi_{11}) \geq r(1/2 - r)(K^2 - \Delta^2)^2$. Indeed, one compares pairs of points (x, x') and (y, y') for $(x, y) \in A_1 \times A_1$ and $(x', y') \in A_1 \times A_2$, therefore $(x - x')^2$ is upper bounded by Δ^2 and $(y - y')^2$ lower bounded by K^2 and the bound above follows after integration against the

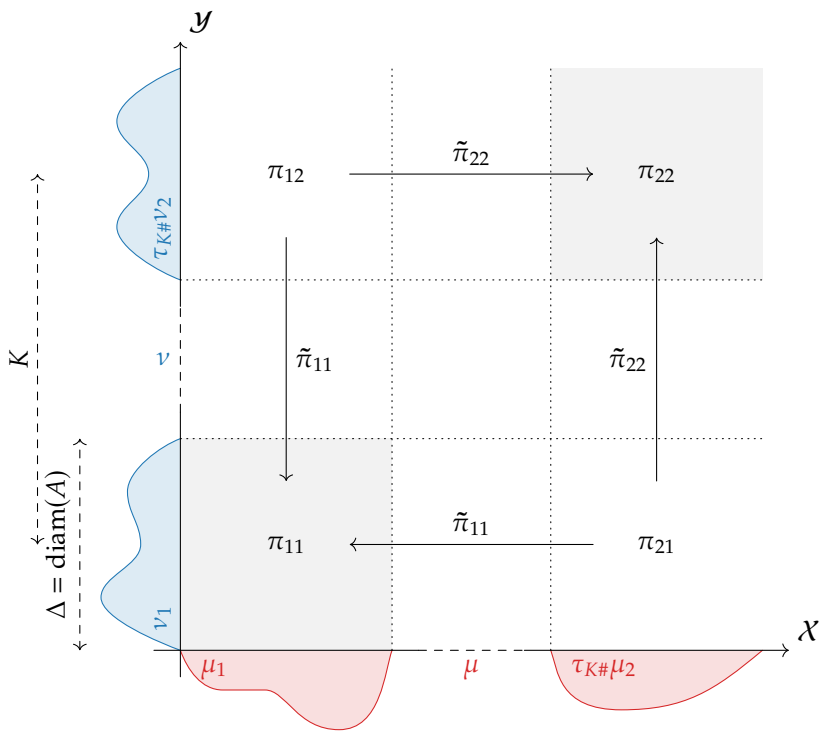


Figure 6: Visual sketch of the proof of Lemma 5.

corresponding measures. The second type is $\mathrm{GW}(\pi_{12}, \pi_{21}) \geq 0$, there are 4 of such terms. We thus have $R \geq 8r(1/2 - r)(K^2 - \Delta^2)^2$. We now upper-bound the competitor. Similarly, one has

$$\mathrm{GW}(\tilde{\pi}, \tilde{\pi}) = \sum_{i,j} \mathrm{GW}(\pi_{ii}, \pi_{jj}) + \tilde{R}$$

where $\tilde{R} = 2\,\mathrm{GW}(\tilde{\pi}_{11}, \pi_{22} + \tilde{\pi}_{22}) + 2\,\mathrm{GW}(\tilde{\pi}_{22}, \pi_{11} + \tilde{\pi}_{11}) + 2\,\mathrm{GW}(\pi_{11}, \tilde{\pi}_{11}) + 2\,\mathrm{GW}(\pi_{22}, \tilde{\pi}_{22})$. The two last terms can be upper bounded by $2r(1/2 - r)\Delta^2$. Indeed, one compares distance squared of couples of points in A_1 to couple of points in A_1, so it is upper bounded by Δ^2. Again by elementary inequalities (see Figure 6), the two first terms can be upper bounded by $r(2K\Delta + \Delta^2)^2$. Note that the total mass of the plan $\pi_{11} + \tilde{\pi}_{11}$ is $1/2$ which explains why $(1/2 - r)$ does not appear. Therefore, the difference between the two values of GW is

$$\mathrm{GW}(\pi, \pi) - \mathrm{GW}(\tilde{\pi}, \tilde{\pi}) \geq r\left(8(1/2 - r)(K^2 - \Delta^2)^2 - 4(1/2 - r)\Delta^2 - 2(2K\Delta + \Delta^2)^2\right). \tag{12}$$

Then, since $1/2 - r \geq 1/4$ the limit in K of the polynomial function on the r.h.s. of Eq. (12) is $+\infty$ uniformly in $r \in [0, \frac{1}{4}]$, and the result follows; there exists $K > 0$ such that the polynomial function above is non-negative, for instance $\max(0, K_0)$ where K_0 is the largest root.

The proof in the case $t > 1/2$ (the other is symmetric) is even simpler since $t - r > t - 1/2$ and consequently, there is no choice in the matching of the two measures; it is determined by the corresponding masses. One can directly apply the argument above. □

We now prove Proposition 15.

Proof of Proposition 15. Thanks to Lemma 5, we know that we can restrict to transportation plans $\pi = \pi_1 + \pi_2$ where, up to flipping ν, we can assume that π_1 is supported on $A \times A$ and π_2 on

$(A+K)\times(A+K)$.[3]

Using again the bilinear form $\mathrm{GW}(\pi,\gamma)$ defined in (11), the objective value reached by any transport plan $\pi=\pi_1+\pi_2$ actually decomposes as

$$\mathrm{GW}(\pi,\pi)=\mathrm{GW}(\pi_1,\pi_1)+2\,\mathrm{GW}(\pi_1,\pi_2)+\mathrm{GW}(\pi_2,\pi_2)\,.$$

Now, assume that we have found $\pi_2^\star$ optimal. Let us minimize in π_1 the resulting quadratic problem:

$$\min_{\pi_1}\ \mathrm{GW}(\pi_1,\pi_1)+2\,\mathrm{GW}(\pi_1,\pi_2^\star)\,.$$

We know that if $\pi_1^\star$ is a minimizer of this quantity, it must also be a solution of the *linear* problem

$$\min_{\pi_1}\ \mathrm{GW}(\pi_1,\pi_1^\star)+\mathrm{GW}(\pi_1,\pi_2^\star)\,.$$

This minimization problem is exactly the optimal transportation problem for the cost

$$c(x,y)\ =\ \int_{A\times A}((x-x')^2-(y-y')^2)^2\,\mathrm{d}\pi_1^\star(x',y')+2\int_{(A+K)^2}((x-x'')^2-(y-y'')^2)^2\,\mathrm{d}\pi_2^\star(x'',y'')\,.$$

Now, using the relation $((x-x'')^2-(y-y'')^2)^2=((x-y)-(x''-y''))^2((x+y)-(x''+y''))^2$, and that $\pi_2^\star$ is a transportation plan between $\tau_{K\#}\mu_2$ and $\tau_{K\#}\nu_2$ so that we can make a change of variable, observe that

$$\begin{aligned}c(x,y)=&\int_{A\times A}((x-x')^2-(y-y')^2)^2\,\mathrm{d}\pi_1^\star(x',y')\\&+\int_{A\times A}((x-y)-(x''-y''))^2((x+y)-(x''+y''+2K))^2\,\mathrm{d}(\tau_{-K},\tau_{-K})_\#\pi_2^\star(x'',y'')\,.\end{aligned}$$

Now, observe that $\partial_{xy}c(x,y)$ is a polynomial function in K,x,y whose dominant term in K is simply $-2K^2$, and recall that A is compact, so that this polynomial function is bounded in x,y. We conclude

$$\partial_{xy}c(x,y)=-2K^2+O(K)<0$$

for K large enough, for all $(x,y)\in A\times A$. The plan $\pi_1^\star$ is optimal for a submodular cost, and by Proposition 12 must be the non-decreasing matching between μ_1 and ν_1. By symmetry, so is $\pi_2^\star$. □

[3]Note: this is where the choice is made, as in the proof of Lemma 5, between the increasing and the non-increasing matchings. Using this convention, the non-decreasing monotone map is shown to be optimal.

Acknowledgements

The authors would like to thank Facundo Mémoli and Gabriel Peyré for their helpful remarks. This work was supported by the Bézout Labex (New Monge Problems), funded by ANR, reference ANR-10-LABX-58.

References

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005. 8, 36

[AKM11] Najma Ahmad, Hwa Kil Kim, and Robert J McCann. Optimal transportation, topology and uniqueness. *Bulletin of Mathematical Sciences*, 1(1):13–32, 2011. 6

[AMJ18] David Alvarez-Melis and Tommi S Jaakkola. Gromov–wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018. 2, 3

[ANN18] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Snowflake universality of wasserstein spaces. In *Annales Scientifiques de l'Ecole Normale Superieure*, volume 51, pages 657–700. Societe Mathematique de France, 2018. 5

[BAMKJ19] Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning generative models across incomparable spaces. *arXiv preprint arXiv:1905.05461*, 2019. 2

[BBM05] Alexander C Berg, Tamara L Berg, and Jitendra Malik. Shape matching and object recognition using low distortion correspondences. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 26–33. IEEE, 2005. 2

[BHS22] Robert Beinert, Cosmas Heiss, and Gabriele Steidl. On assignment problems related to gromov–wasserstein distances on the real line. *arXiv preprint arXiv:2205.09006*, 2022. 2, 7, 20, 21, 22, 23, 24

[Bre87] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987. 2, 5

[BTBD20] Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *International Conference on Machine Learning*, pages 950–959. PMLR, 2020. 20

[Car12] Guillaume Carlier. Optimal transportation and economic applications. *Lecture Notes*, 2012. 18

[Chi07] David Chiron. On the definitions of sobolev and bv spaces into singular spaces and the trace problem. *Communications in Contemporary Mathematics*, 9(04):473–513, 2007. 35

[CMN10] Pierre-André Chiappori, Robert J McCann, and Lars P Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, 42(2):317–354, 2010. 6, 22

[DPM22] Nicoló De Ponti and Andrea Mondino. Entropy-transport distances between unbalanced metric measure spaces. *Probability Theory and Related Fields*, 184(1):159–208, 2022. 2

[DSC+22] Pınar Demetçi, Rebecca Santorella, Manav Chakravarthy, Bjorn Sandstede, and Ritambhara Singh. Scotv2: Single-cell multiomic alignment with disproportionate cell-type representation. *Journal of Computational Biology*, 29(11):1213–1228, 2022. 2

[DSS+20] Pınar Demetçi, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Gromov–wasserstein optimal transport to align single-cell multi-omics data. *BioRxiv*, 2020. 2

[FGM10] Joaquin Fontbona, Hélène Guérin, and Sylvie Méléard. Measurability of optimal transportation and strong coupling of martingale measures. *Electronic communications in probability*, 15:124–133, 2010. 12, 14, 33, 34, 35

[FH74] Jacques Faraut and Khélifa Harzallah. Distances hilbertiennes invariantes sur un espace homogene. In *Annales de l'institut Fourier*, volume 24, pages 171–217, 1974. 5

[GJB19] Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1880–1890. PMLR, 2019. 2

[GKPS99] Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999. 2, 3

[HGL21] YoonHaeng Hur, Wenxuan Guo, and Tengyuan Liang. Reversible gromov-monge sampler for simulation-based inference, 2021. 7

[KB57] Tjalling C Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society*, pages 53–76, 1957. 2, 3

[Kon76] Hiroshi Konno. Maximization of a convex quadratic function under linear constraints. *Mathematical programming*, 11(1):117–127, 1976. 4

[Lyo13] Russell Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284–3305, 2013. 5

[McC01] Robert J McCann. Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001. 2, 5

[McC12] Robert J McCann. A glimpse into the differential topology and geometry of optimal transport. *arXiv preprint arXiv:1207.1867*, 2012. 6, 18

[Mém07] Facundo Mémoli. On the use of gromov–hausdorff distances for shape comparison. 2007. 2, 3, 4

[Mém08] Facundo Mémoli. Gromov–hausdorff distances in euclidean spaces. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8. IEEE, 2008. 2

[Mém11] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011. 2

[MG11] Robert J McCann and Nestor Guillen. Five lectures on optimal transportation: geometry, regularity and applications. *Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal*, pages 145–180, 2011. 6

[ML18] Haggai Maron and Yaron Lipman. (probably) concave graph matching. *Advances in Neural Information Processing Systems*, 31, 2018. 3

[MN22] Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022. 2, 7

[Moa16] Abbas Moameni. A characterization for solutions of the monge-kantorovich mass transport problem. *Mathematische Annalen*, 365(3):1279–1304, 2016. 6

[MPW12] Robert J McCann, Brendan Pass, and Micah Warren. Rectifiability of optimal transportation plans. *Canadian Journal of Mathematics*, 64(4):924–934, 2012. 7

[MR20] Abbas Moameni and Ludovic Rifford. Uniquely minimizing costs for the kantorovitch problem. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 29, pages 507–563, 2020. 7

[MS04] Facundo Mémoli and Guillermo Sapiro. Comparing point clouds. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, 2004. 2

[MS05] Facundo Mémoli and Guillermo Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5(3):313–347, 2005. 2

[MSW21] Facundo Mémoli, Zane Smith, and Zhengchao Wan. The gromov–hausdorff distance between ultrametric spaces: its structure and computation. *arXiv preprint arXiv:2110.03136*, 2021. 2

[MTOL20] Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020. 2

[PC19] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 5

[PGM+19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 20

[RVFC20] Ivegen Redko, Titouan Vayer, Rémi Flamary, and Nicolas Courty. Co-optimal transport. *Advances in Neural Information Processing Systems*, 33:17559–17570, 2020. 2

[RW09] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009. 14, 32, 33

[San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015. 2, 5, 11, 18

[Sch38] Isaac J Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44:522–536, 1938. 5

[SDD21] Antoine Salmona, Julie Delon, and Agnès Desolneux. Gromov–wasserstein distances between gaussian distributions. *arXiv preprint arXiv:2104.07970*, 2021. 7

[Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006. 2

[Stu12] Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012. 2, 3, 7

[SVP21] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. *Advances in Neural Information Processing Systems*, 34:8766–8779, 2021. 2, 4

[Ume88] Shinji Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence*, 10(5):695–703, 1988. 2

[Vay20] Titouan Vayer. A contribution to optimal transport on incomparable spaces. *arXiv preprint arXiv:2011.04447*, 2020. 2, 3, 7, 8, 15, 16, 20

[VdV00] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. 35

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2008. 5, 6, 7, 34

[XLH+20] Hongteng Xu, Dixin Luo, Ricardo Henao, Svati Shah, and Lawrence Carin. Learning autoencoders with relational regularization. *arXiv preprint arXiv:2002.02913*, 2020. 2

[ZDlT15] Feng Zhou and Fernando De la Torre. Factorized graph matching. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1774–1789, 2015. 2

A Appendix

A.1 Proofs of Lemmas 3 and 4

Proof of Lemma 3. Remark that the continuity of ψ_1 and ψ_2 and their inverse ensures their measurability. We have the following equalities:

$$\begin{aligned}\operatorname*{arg\,min}_{\pi\in\Pi(\mu,\nu)} \int \tilde{c}(\psi_1(x),\psi_2(y))\,\mathrm{d}\pi(x,y) &= \operatorname*{arg\,min}_{\pi\in\Pi(\mu,\nu)} \int \tilde{c}(u,v)\,\mathrm{d}(\psi_1,\psi_2)_\#\pi(u,v)\\ &= (\psi_1^{-1},\psi_2^{-1})_\# \operatorname*{arg\,min}_{\tilde{\pi}\in\Pi(\psi_{1\#}\mu,\psi_{2\#}\nu)} \int \tilde{c}(u,v)\,\mathrm{d}\tilde{\pi}(u,v)\end{aligned}$$

since the mapping $(\psi_1^{-1},\psi_2^{-1})_\#$ is a one-to-one correspondence from $\Pi(\psi_{1\#}\mu,\psi_{2\#}\nu)$ to $\Pi(\mu,\nu)$ by bijectivity of ψ_1 and ψ_2. This bijectivity ensures that any optimal deterministic transport plan $\tilde{\pi}^\star$ between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ induces an optimal deterministic transport plan $\pi^\star$ between μ and ν, and *vice versa*. Writing $\tilde{\pi}^\star = (\mathrm{id},T)_\#(\psi_{1\#}\mu)$, this plan $\pi^\star$ is given by

$$\begin{aligned}\pi^\star &= (\psi_1^{-1},\psi_2^{-1})_\#\tilde{\pi}^\star\\ &= (\psi_1^{-1},\psi_2^{-1})_\#(\mathrm{id},T)_\#\psi_{1\#}\mu\\ &= (\mathrm{id},\psi_2^{-1}\circ T\circ\psi_1)_\#\mu\,.\end{aligned}$$

□

Proof of Lemma 4. As $\psi_{1\#}\mu$ has a density w.r.t. the Lebesgue measure since ψ_1 is a diffeomorphism and $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ have compact support, Brenier's theorem states that there exists a unique optimal transport plan between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ and that it is induced by a map $\nabla\varphi$, where φ is a convex function. Using Lemma 3 then gives the result. □

Remark 11 (Discussion on the hypothesis of Lemma 4). In the proof of Lemma 4, we only needed (i) ψ_1, ψ_2 and their inverse to be measurable, (ii) $\psi_{1\#}\mu$ to have a density w.r.t. Lebesgue, and (iii) $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ to have compact support. Imposing ψ_1 to be a diffeomorphism and ψ_2 to be a homeomorphism ensures both (i) and (ii) and is natural to expect.

A.2 Measurable selection of maps in the manifold setting

A.2.1 Measurability of set-valued maps. Let X, U be two topological spaces, and let $\mathcal{B}$ denote the Borel σ-algebra on X. A set-valued map S is a map from X to $P(U)$ (the set of subsets of U). This will be denoted by $S : X \rightrightarrows U$. The idea is to introduce notations which are consistent with the case where $S(x) = \{u\}$ for all x in X, where we want to retrieve the standard case of maps $X \to U$. Definitions are taken from [RW09], where measurability is studied when $U = \mathbb{R}^n$. Most results and proofs adapt to a more general setting—in particular when U is a complete Riemannian manifold M, which we shall assume in the following. For the sake of completeness, we provide all the proofs, and highlight those that require specific care when replacing $\mathbb{R}^n$ by a manifold.

Of importance for our proofs, we define:

- The *pre-image* of a set $B \subset U$ is given by $S^{-1}(B) = \{x \in X \mid S(x) \cap B \neq \varnothing\}$.
- The *domain* of S is $S^{-1}(U)$, that is $\{x \in X \mid S(x) \neq \varnothing\}$.

We will often use the following relation: if a set A can be written as $A = \bigcup A_k$, then $S^{-1}(A) = \bigcup S^{-1}(A_k)$. Indeed, $x \in S^{-1}(A) \Leftrightarrow S(x) \cap A \neq \varnothing \Leftrightarrow \exists k,\ S(x) \cap A_k \neq \varnothing \Leftrightarrow \exists k,\ x \in S^{-1}(A_k) \Leftrightarrow x \in \bigcup_k S^{-1}(A_k)$.

A set-valued map $S : X \rightrightarrows U$ is said to be *measurable* if, for any open set $O \subset U$,

$$S^{-1}(O) \in \mathcal{B}.$$

Note that if S is measurable (as a set-valued map), then its domain must be measurable as well (as an element of $\mathcal{A}$). We say that $S : X \rightrightarrows U$ is *closed-valued* if $S(x)$ is a closed subset of U for all $x \in X$.

Proposition 16 (Theorem 14.3.c in [RW09])**.** A closed-valued map $S : X \rightrightarrows U$ is measurable if and only if $S^{-1}(B) \in \mathcal{B}$ holds either:
(a) for all $B \subset U$ open (the definition);
(b) for all $B \subset U$ compact;
(c) for all $B \subset U$ closed.

Proof of Proposition 16.
- (a) $\Rightarrow$ (b): For a compact $B \subset U$, let $B_k = \{x \in U \mid d(x, B) < k^{-1}\}$, $k \geq 0$ (that is open). Note that $x \in S^{-1}(B) \Leftrightarrow S(x) \cap B \neq \varnothing \Leftrightarrow S(x) \cap B_k \neq \varnothing$ for all k because $S(x)$ is a closed set. Hence $S^{-1}(B) = \bigcap_k S^{-1}(B_k)$. All the $S^{-1}(B_k)$ are measurable, so is $S^{-1}(B)$ as a countable intersection of measurable sets.
- (b) $\Rightarrow$ (a): Fix O an open set of U. As we assume U to be a complete separable Riemannian manifold, O can be written as a countable union of compact balls: $O = \bigcup_n \overline{B(x_n, r_n)}$.
- (b) $\Rightarrow$ (c): Immediate.
- (c) $\Rightarrow$ (b): A closed set B can be obtained as a countable union of compact sets by letting $B = \bigcup_n B \cap \overline{B(x_0, n)}$ for some x_0. Hence $S^{-1}(B) = \bigcup_n S^{-1}(B \cap \overline{B(x_0, n)})$ is in $\mathcal{B}$. □

Now, we introduce a proposition on operations that preserve measurability of closed-set valued maps. The proof requires adaptation from the one of [RW09] because the latter uses explicitly the fact that one can compute Minkowski sums of sets (which may not make sense on a manifold).

Proposition 17 (Proposition 14.11 in [RW09], adapted to the manifold case)**.** Let S_1 and $S_2 : X \rightrightarrows U$ be two measurable closed-set valued maps. Then
- $P : x \mapsto S_1(x) \times S_2(x)$ is measurable as a closed-valued map in $U \times U$ (equipped with the product topology).
- $Q : x \mapsto S_1(x) \cap S_2(x)$ is measurable.

Proof. The first point can be proved in the same spirit as the proof proposed by Rockafellar and Wets. Namely, let O' be an open set in $U \times U$. By definition of the product topology, O' can be obtained as $\bigcup_n O_1^{(n)} \times O_2^{(n)}$ where $O_1^{(n)}$ and $O_2^{(n)}$ are open sets in U. Then $P^{-1}(O') = \bigcup_n P^{-1}(O_1^{(n)} \times O_2^{(n)})$. Now, observe that $P^{-1}(A \times B) = \{x \mid S_1(x) \times S_2(x) \in A \times B\} = \{x \mid S_1(x) \in A \text{ and } S_2(x) \in B\} = S_1^{-1}(A) \cap S_2^{-1}(B)$, so that finally, $P^{-1}(O') = \bigcup_n S_1^{-1}(O_1^{(n)}) \cap S_2^{-1}(O_2^{(n)})$ that is measurable as a countable union of (finite) intersection of measurable sets (given that S_1, S_2 are measurable). Note that this does not require S_1, S_2 to be closed-valued.

Now, let us focus on the second point, that requires more attention. Thanks to the previous proposition, it is sufficient to show that $Q^{-1}(C) \in \mathcal{B}$ for any compact set $C \subset U$. In [RW09], this is done by writing $Q^{-1}(C) = \{x \mid S_1(x) \cap S_2(x) \cap C \neq \varnothing\} = \{x \mid R_1(x) \cap R_2(x) \neq \varnothing\} = \{x \mid 0 \in (R_1(x) - R_2(x))\} = (R_1 - R_2)^{-1}(\{0\})$, where $R_j(x) = S_j(x) \cap C$ (that is also closed valued), and using the fact that the (Minkowski) difference of measurable closed-valued maps is measurable as well [RW09, Prop. 14.11.c].

To adapt this idea (we cannot consider Minkowski difference in our setting), we introduce the diagonal $\Delta = \{(u, u) \mid u \in U\} \subset U \times U$. Now, observe that $R_1(x) \cap R_2(x) \neq \varnothing \Leftrightarrow (R_1(x) \times R_2(x)) \cap \Delta \neq \varnothing$, that is $x \in R^{-1}(\Delta)$, where $R(x) = R_1(x) \times R_2(x)$. Since the maps R_1 and R_2 are measurable closed-valued maps (inherited from S_1, S_2), so is R according to the previous point. And since Δ is closed, $R^{-1}(\Delta) = Q^{-1}(C)$ is measurable. □

A.2.2 Proof of Proposition 11. The proof is essentially an adaptation of the one of [FGM10], with additional care required due to the fact that we do not have access to a linear structure on the manifold M. It relies on measurability of set-valued maps (see [RW09, Ch. 5 and 14] and Appendix A.2.1 for a summary).

In the following, we consider a partition $(A_{n,k})_n$ of M made of cells with diameter $\leq 2^{-k}$ and such that $(A_{n,j})_n$ is a refinement of $(A_{n,k})_n$ in sense that each $A_{n,k}$ is itself partitioned by some $A_{n',j}$. The crucial point regarding measurability is the following proposition.

Proposition 18. The set
$$B_{n,k} = \{(u,x) \mid T_u(x) \in A_{n,k}\}.$$
is measurable.

Its proof relies on a core lemma:

Lemma 6. Let $F \subset M$ be a closed set. Then the set
$$B_F = \{(u,x) \mid T_u(x) \in F\}$$
is measurable.

The key will be to identify this set as the domain of a measurable set-valued map, see Appendix A.2.1.

Proof of Lemma 6. Observe that $B_F = \{(u,x) \mid (\{x\} \times F) \cap \mathrm{gph}(T_u) \neq \varnothing\}$, where $\mathrm{gph}(T_u) = \{(x, T_u(x)) \mid x \in M\}$ denotes the topological closure of the graph of the optimal transport map T_u that pushes μ_u onto ν_u. Let $S_1 : (u,x) \mapsto \{x\} \times F$ and $S_2 : (u,x) \mapsto \mathrm{gph}(T_u)$, so that $B_F = \mathrm{dom}(S)$, where $S(x) = S_1(x) \cap S_2(x)$. According to Proposition 17, given that S_1 and S_2 are closed-valued, if they are measurable, so is S, and so is B_F as the domain of a measurable map. The measurability of these two maps can be easily adapted from the work of [FGM10], we give details for the sake of completeness.

Measurability of S_1: Let $O \subset M \times M$ be open. Observe that $S_1^{-1}(O) = \{(u,x) \mid \{x\} \times F \cap O \neq \varnothing\} = B \times \{x \mid \{x\} \times F \cap O \neq \varnothing\}$. Fix $z \in F$. We have $\varepsilon > 0$ such that $B(x,\varepsilon) \times \{z\} \subset O$ (since O is open), and thus $B(x,\varepsilon) \times F \cap O \neq \varnothing$, proving that there is a neighborhood of x included in $\{x \mid \{x\} \times F \cap O \neq \varnothing\}$ which is thus open, making in turn $S_1^{-1}(O)$ open (thus measurable), hence the measurability of S_1.

Measurability of S_2: Given that $u \mapsto (\mu_u, \nu_u)$ is measurable by assumption, and that measurability is preserved by composition, we want to show that (i) the map $S : (\mu,\nu) \mapsto \Pi^\star(\mu,\nu)$ (the set of optimal transport plans between μ and ν for the quadratic cost on M) is measurable and (ii) the map $\mathfrak{U} : \pi \in \mathcal{P}(M^2) \mapsto \mathrm{supp}\,\pi$ satisfies $\mathfrak{U}^{-1}(O)$ is open for any open set $O \subset \mathcal{P}(M^2)$. From these two points, we get that $(U \circ S)^{-1}(O)$ is measurable, thus the measurability of S_2.

To get (i), observe first that S is closed-valued, so that it is sufficient to prove that $S^{-1}(C)$ is measurable for any closed set $C \subset \mathcal{P}(M^2)$ according to Proposition 16. Let $C \subset \mathcal{P}(M^2)$ be closed. Then, $S^{-1}(C) = \{(\mu,\nu) \mid \Pi^\star(\mu,\nu) \cap C \neq \varnothing\}$, and consider a sequence $(\mu_n,\nu_n)_n$ in $S^{-1}(C)$ that converges to (μ,ν) for the weak topology. Let $\pi_n \in \Pi^\star(\mu_n,\nu_n) \cap C$. According to [Vil08, Thm. 5.20], $(\pi_n)_n$ admits a weak limit π in $\Pi^\star(\mu,\nu)$, but also since C is closed, $\pi \in C$, so that $(\mu,\nu) \in S^{-1}(C)$ that is closed (hence measurable), proving the measurability of S.

(ii) simply follows from the fact that $\mathfrak{U}^{-1}(O) = \{\pi \mid \mathrm{supp}\,\pi \cap O \neq \varnothing\} = \{\pi \mid \pi(O) > 0\}$ that is open. Indeed, the Portmanteau theorem gives that if $\pi_n \to \pi$ (weakly) and $\pi_n(O) = 0$, then $0 = \liminf \pi_n(O) \geq \pi(O) \geq 0$, so $\pi(O) = 0$. The complementary set of $U^{-1}(O)$ is closed, that is $\mathfrak{U}^{-1}(O)$ is open. □

Proof of Proposition 18. Using that $A_{n,k}$ can be inner-approximated by a sequence of closed set $F_j \subset A_{n,k}$, we obtain a sequence of sets $(B_{F_j})_j$ such that $\bigcup_j B_{F_j} = A_{n,k}$. By Lemma 6, the $(B_{F_j})_j$ are measurable, so is $A_{n,k}$ as the (countable) union of measurable sets. □

We can now prove our main proposition regarding measurability.

Proof of Proposition 11. Recall that we assume that $M = \bigsqcup_n A_{n,k}$. For each n,k, select (in a measurable way) a $a_{n,k}$ in $A_{n,k}$. Then, define the map
$$T^{(k)} : (u,x) \mapsto a_{n,k}, \quad \text{such that } T_u(x) \in A_{n,k}.$$
This map is measurable. Indeed, for any open $O \subset M$, $T^{(k),-1}(O) = \{(u,x) \mid T_u(x) \in A_{n,k}\} = B_{n,k}$ that is measurable according to Proposition 18.

Now, for two maps $f, g : B \times M \to M$, let D_1 denotes the natural L_1 distance on M, that is
$$D_1(f,g) = \int_B \int_M d(f(u,x), g(u,x))\, \mathrm{d}\mu_u(x)\, \mathrm{d}m(u)\,.$$

This yields a complete metric space [Chi07], and we can observe that $(T^{(k)})_k$ is a Cauchy sequence for this distance. Indeed, for $k \leq j$ two integers, recall that we assume that $(A_{n,j})_n$ is a refinement of $(A_{n,k})_n$, yielding

$$\begin{aligned}
D_1(T^{(k)}, T^{(j)}) &= \int_B \int_M d(T^{(k)}(u,x), T^{(j)}(u,x))\, \mathrm{d}\mu_u(x)\, \mathrm{d}m(u) \\
&= \int_B \int_M \sum_n \sum_{n' : A_{n',j} \subset A_{n,k}} 1_{B_{n',j}}(u,x) \cdot d(a_{n,k}, a_{n',j})\, \mathrm{d}\mu_u(x)\, \mathrm{d}m(u) \\
&= \int_B \int_M \sum_n \sum_{n' : A_{n',j} \subset A_{n,k}} d(a_{n,k}, a_{n',j})\, \mathrm{d}\nu_u(A_{n',j})\, \mathrm{d}m(u) \\
&\leq 2^{-k}
\end{aligned}$$

where we use that for all u, $\int_{x \in M} 1_{B_{n',j}}(u,x)\, \mathrm{d}\mu_u(x) = \nu_u(A_{n',j})$ by construction (recall that $(u,x) \in B_{n',j} \Leftrightarrow T_u(x) \in A_{n',j} \Leftrightarrow x \in \mu_u(T_u^{-1}(A_{n',j})) = T_u \# \mu_u(A_{n',j})$ and T_u transports μ_u onto ν_u), and then that the diameter of the partition $A_{n,k}$ is less than or equal to 2^{-k} and that ν_u and m are probability measures.

Now, let T denote the limit of $(T^{(k)})_k$ (that is measurable). It remains to show that $T(u,x) = T_u(x)$, m-a.e. This can be obtained by proving that

$$\int g(x) f(T(x,u))\, \mathrm{d}\mu_u(x) = \int g(x) f(T_u(x))\, \mathrm{d}\mu_u(x)\,,$$

for any pair $f, g : M \to \mathbb{R}$ of bounded Lipschitz-continuous functions [VdV00, Lemma 2.24].

As in [FGM10], let $\|f\| := \sup_{x \neq y} \frac{|f(x)-f(y)|}{d(x,y)} + \sup_x |f(x)|$. The difference between these two terms can be bounded using the partition $(A_{n,k})_n$. We have for m-a.e. u:

$$\begin{aligned}
&\left| \int g(x) f(T_u(x))\, \mathrm{d}\mu_u(x) - \int g(x) f(T(u,x))\, \mathrm{d}\mu_u(x) \right| \\
\leq &\left| \int g(x) f(T_u(x))\, \mathrm{d}\mu_u(x) - \int g(x) f(T^{(k)}(u,x))\, \mathrm{d}\mu_u(x) \right| + \|g\| \|f\| \int d\left(T^{(k)}(u,x), T(u,x)\right)\, \mathrm{d}\mu_u(x)\,.
\end{aligned}$$

Since $T^{(k)} \to T$ in D_1, it implies that up to a subsequence, $\int_x d(T^{(k)}(u,x), T(u,x))\, \mathrm{d}\mu_u(x) \to 0$ as $k \to \infty$ for m-a.e. u.

To treat the first term and show that it goes to 0 as $k \to \infty$ for a subset of B with full m-measure, we write for m-a.e. u:

$$\begin{aligned}
&\left| \int g(x) f(T_u(x))\, \mathrm{d}\mu_u(x) - \int g(x) f(T^{(k)}(u,x))\, \mathrm{d}\mu_u(x) \right| \\
\leq &\int |g(x)| \left| f(T_u(x)) - f(T^{(k)}(u,x)) \right|\, \mathrm{d}\mu_u(x) \\
\leq &\|g\| \|f\| \int d(T_u(x), T^{(k)}(u,x))\, \mathrm{d}\mu_u(x) \\
\leq &\|g\| \|f\|\, 2^{-k} \sum_n \nu_u(A_{n,k}) \to 0\,.
\end{aligned}$$

This concludes the proof. □

A.3 Measure disintegration

The following definition and theorem are taken from [AGS05].

Definition 5 (Measure disintegration)**.** *Let $\mathcal{X}$ and $\mathcal{Z}$ be two Radon spaces, $\mu \in P(\mathcal{X})$ and $\varphi : \mathcal{X} \to \mathcal{Z}$ a Borel-measurable function. A family of probability measures $\{\mu_u\}_{u\in\mathcal{Z}} \subset P(\mathcal{X})$ is a* disintegration *of μ by φ if:*

(i) the function $u \mapsto \mu_u$ is Borel-measurable;
(ii) μ_u lives on the fiber $\varphi^{-1}(\{u\})$: for $\varphi_\#\mu$-a.e. $u \in \mathcal{Z}$,

$$\mu_u(\mathcal{X} \setminus \varphi^{-1}(\{u\})) = 0\,,$$

and so $\mu_u(B) = \mu_u(B \cap \varphi^{-1}(\{u\}))$ for any Borel $B \subset \mathcal{X}$;
(iii) for every measurable function $f : \mathcal{X} \to [0, \infty]$,

$$\int_{\mathcal{X}} f(x)\,\mathrm{d}\mu(x) = \int_{\mathcal{Z}} \left(\int_{\varphi^{-1}(\{u\})} f(x)\,\mathrm{d}\mu_u(x) \right) \mathrm{d}(\varphi_\#\mu)(u)\,.$$

In particular, for any Borel $B \subset \mathcal{X}$, taking f to be the indicator function of B,

$$\mu(B) = \int_{\mathcal{Z}} \mu_u(B)\,\mathrm{d}(\varphi_\#\mu)(u)\,.$$

Theorem 6 (Disintegration theorem)**.** Let $\mathcal{X}$ and $\mathcal{Z}$ be two Radon spaces, $\mu \in P(\mathcal{X})$ and $\varphi : \mathcal{X} \to \mathcal{Z}$ a Borel-measurable function. There exists a $\varphi_\#\mu$-a.e. uniquely determined family of probability measures $\{\mu_u\}_{u\in\mathcal{Z}} \subset P(\mathcal{X})$ that provides a disintegration of μ by φ.